Maria Rosana Ponisio, MD
11339 Mosley Forest Dr
St. Louis, MO 63141

October 25, 2021

Corey J. Smith
Senior Litigation Counsel
Tax Division

Re: Brockman, Robert ████████████1941)

As you requested, I have reviewed the FDG PET/CT scan dated (03/12/2021), amyloid PET/CT scan (07/28/2021), and FDG PET/CT (08/24/2021) of the individual named above. Qualitatively and quantitatively interpretation using the MIM software (MIMneuro) was performed. The PET images were fused with the brain MRI examination dated 07/30/2021 (axial MPR).

FDG PET/CT scan date 03/12/2021 demonstrates abnormal moderate to markedly decreased metabolic activity in the cingulate gyrus and bilateral precuneus. Mild hypometabolism was noted in the right frontoparietal lobes. There is marked decreased FDG avidity in the bilateral caudate nucleus with no significant decreased metabolic activity seen in the frontal or occipital lobes. Using MIMneuro, a quantitative analysis was performed. A statistically significant decreased tracer accumulation ($Z <= -2$) is seen in the caudate nuclei, anterior cingulate, posterior cingulate, and precuneus.

FDG PET/CT dated 08/24/2021 demonstrates abnormal markedly decreased metabolic activity in the cingulate gyrus, more pronounced in the posterior cingulate, and moderate in the bilateral precuneus. Mild hypometabolism is noted in the right frontoparietal lobes. In addition, there is a marked decreased FDG avidity in the bilateral caudate nucleus. No significant decreased metabolic activity is seen in the frontal or occipital lobes. The cerebellum demonstrates normal

Ex. 86 p. 1 of 3

metabolic activity. A statistically significant decreased tracer accumulation (Z <= -2) is seen in the caudate nuclei, posterior cingulate, left amygdala, anterior cingulate, precuneus, and superior frontal gyrus.

The brain amyloid PET/CT (F-18 florbetapir, amyloid) dated 07/28/2021 showed an increased grey matter cortical activity in the bilateral temporal, frontal, and parietal lobes. There is also increased tracer accumulation in the bilateral precuneus and posterior cingulate gyrus. Statistical significant tracer deposition in the cerebral cortex (Z score <= -2) was observed in the precuneus, posterior cingulate, anterior cingulate, bilateral lateral temporal lobes, superior parietal lobes, and inferior frontal gyrus. Normal activity was noted in the cerebellum

Combined Scans Interpretation:

Both FDG PET/CT shows metabolic findings most consistent with early Alzheimer dementia in the correct clinical setting. This interpretation is based on the abnormal metabolic pattern seen on the visual and quantitative analysis, which shows decreased metabolic activity involving the precuneus and the cingulate gyrus. These areas are affected in the early stages of Alzheimer's disease. In addition, there was mild to moderate hypometabolism in the parietal and frontal lobes; also, common regions compromise in Alzheimer's disease. The decreased metabolic activity in these affected areas was slightly more pronounced in the second FDG scan.

The absence of significant decreased metabolic activity in the frontal and anterior temporal lobes excludes frontotemporal dementia as a probable diagnosis. Furthermore, normal metabolic activity in the occipital lobe makes dementia of Lewis's body unlikely. There was no pattern of hypometabolism visualized to suggest vascular dementia.

The brain amyloid PET/CT (F-18 florbetapir, amyloid) scan (07/28/2021) shows abnormal diffuse tracer deposition in the grey matter consistent with a positive amyloid-PET study indicating

moderate to frequent neuritic plaques, which is the amount present in patients with Alzheimer's disease. However, a positive amyloid-PET study does not establish the diagnosis of Alzheimer's disease in isolation since moderate to frequent neuritic plaques can also be present in patients with other neurological conditions and older people with normal cognition.

The findings were not consistent with a negative amyloid-PET study. A negative amyloid-PET scan does not show significant tracer accumulation in the grey matter, indicating sparse to no neuritic plaques, and is inconsistent with a diagnosis of Alzheimer's disease at the time of the study.

In conclusion, given the pattern of metabolic abnormalities seen on FDG/PET-CT and amyloid scans, the findings are most consistent with early dementia in the correct clinical setting.

Maria Rosana Ponisio, MD

U.S. Department of Justice

Tax Division

Washington, D.C. 20530

REZ:LJW:CJSmith
5-11-23920
20162011033

October 2, 2018

BY EMAIL
Scott S. Balber
Herbert Smith Freehills
450 Lexington Avenue 14th Floor
New York, NY 10017

Re: Evatt Tamine
       Statutory Immunity - Agreement

Dear Mr. Balber:

    In accordance with our prior phone conversations, emails, and attorney proffer conducted by you and members of your firm on Thursday September 27, 2018, the government has agreed to grant your client, Evatt Tamine, statutory immunity pursuant to Title 18 U.S.C. § 6003. It is the government's understanding that pursuant to this grant of immunity, Mr. Tamine has agreed to continued long-term cooperation with the government's investigation of Robert Smith, Robert Brockman, and other related persons and entities. Mr. Tamine's cooperation to include: meeting with investigators and prosecutors as requested; providing truthful statements and testimony regarding relevant facts known to him; being available to appear and testify before the Grand Jury, at trial, and at any other relevant court proceeding in the United States; and providing relevant documents within his custody and control. Mr. Tamine understands that a failure to provide truthful testimony, or to otherwise provide false information to the government, can result in his prosecution for perjury or obstruction of justice, and that his grant of immunity does not prevent any such prosecution.
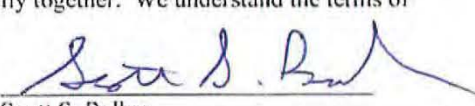
Sincerely,

s/COREY J. SMITH
Senior Litigation Counsel

We have read this letter, and have reviewed it carefully together. We understand the terms of this letter, and we agree to these terms.

_____
Evatt Tamine
Dated: 2 October 2018

_____
Scott S. Balber
Counsel for Evatt Tamine
Dated: October 2, 2018

PARK DIETZ & ASSOCIATES, INC.
Forensic Experts

Administrative Offices

2906 Lafayette
Newport Beach, CA 92663
Tel: 949-723-2211
Fax: 949-723-2212
Email: expert@parkdietzassociates.com
Website: www.parkdietzassociates.com

- **Forensic Psychiatry**
- **Forensic Psychology**
- **Forensic Pathology**
- **Forensic Neurology**
- **Forensic Social Work**
- **Child Welfare**
- **Criminology**
- **Security**

November 4, 2021

Corey J. Smith
Senior Litigation Counsel
U.S. Department of Justice, Tax Division
150 M Street NE, Rm 2.208
Washington, DC 20002

Re: United States v. Robert T. Brockman

Dear Mr. Smith:

This report supplements my reports of June 21, 2021, and October 29, 2021, regarding the competence to stand trial of Mr. Robert Brockman. New information received since that report is listed below, following which I resolve issues left unresolved in my report of October 29, 2021.

NEWLY RECEIVED INFORMATION

Note that the defense experts' supplementary reports cited each other's supplementary reports. Due to the quick turn-around time, the government experts did not exchange reports in advance of the deadline. I first saw Dr. Darby's and Dr. Denney's 10/29/21 reports after submitting my own.

Received 10/29/21 after submission of my 10/29/21 report:

Report of R. Ryan Darby, M.D., 10/29/21

Report of Robert L. Denney, Psy.D., 10/29/21, and test data

Report of Marc E. Agronin, M.D., 10/29/21

Report of Thomas Guilmette, Ph.D., 10/29/21

Report of Christopher T. Whitlow, M.D., Ph.D., M.H.A., 10/29/21

Thomas Wisniewski, M.D., 10/29/21

Received 11/2/21:

Videotape and transcript of Dr. Guilmette's interview of Mr. Brockman, 10/2/21

Testing data from Dr. Guilmette's 10/2/21 testing of Mr. Brockman

Videotape and transcript of Dr. Agronin's interview of Mr. Brockman, 10/3/21

Videotape and transcript of Dr. Wisniewski's interview of Mr. Brockman, 10/17/21

Notes of Dr. Agronin's interview of Dorothy Brockman, 10/3/21

Notes of Dr. Agronin's interview of Frank Gutierrez, 10/3/21

RESOLUTION OF ISSUES

The newly received materials contain nothing that challenges the previously articulated bases of my concerns that Mr. Brockman has continued to exaggerate the degree of his cognitive deficits. On the contrary, they contain evidence demonstrating that Mr. Brockman continues to exaggerate the degree of his impairment.

At the time of my 10/29/21 report, I was uncertain about Mr. Brockman's current memory function, which was the sole basis on which I was uncertain about his trial competence. Dr. Denney's analysis of Mr. Brockman's most recent neuropsychological test results and comparisons with earlier test results provide clear indications that Mr. Brockman continues to exaggerate his degree of impairment. As Dr. Denney wrote based on his detailed analysis, "Mr. Brockman has repeatedly failed validity testing across three different examinations and produced unbelievably severe and inconsistent results."

Moreover, I interpret Dr. Darby's findings as indicating that the seemingly rapid decline in Mr. Brockman's cognitive impairment since May 2021—as displayed in both interview performance and neuropsychological test performance—is inconsistent with the brain imaging studies and the natural history of Parkinson's disease and Alzheimer's disease.

Videotapes of the October interviews also demonstrate significant inconsistencies in Mr. Brockman's apparent level of cognitive impairment, varying from intact thinking skills and preserved memory to an appearance of marked confusion and extreme memory errors. For example, even within the most recent interview by Dr. Denney and me, Mr. Brockman recalled my name and that he didn't like some of our interaction in May, could spell "world" backwards, knew there had been a riot on January 6th, knew that border resources were being overrun, showed sufficient knowledge of trial participants and roles, and remembered the context of the evaluation sufficiently to avoid answering questions about the charges against him. Yet in the same interview he maintained that 20 minus three was 27, initially estimated he'd been married for 35-36 years, estimated he'd been hospitalized

2

overnight 10-12 times in 2021, initially said there'd been no offers on his house, guessed he'd had one hospitalization for urosepsis, said he thought Dr. Denney and I were interviewing him for a research firm hired by a competitor of Reynolds and Reynolds, and said the year was 2051.

Mr. Brockman most likely suffers from mild cognitive impairment or mild dementia, but he continues to malinger neurocognitive dysfunction. His continuing exaggeration and invalid test performance make it impossible to validly <u>measure</u> his memory function. Nonetheless, if Mr. Brockman were so severely demented as to be incompetent to stand trial, he would not be aware that he would benefit from appearing severely demented, would not remember to exaggerate, and would not remember how to exaggerate as well as he does.

Finally, it remains the case that none of the defense expert reports received for review adequately addresses Mr. Brockman's motivation to avoid prosecution, the marked discrepancy between his early cognitive test results and his contemporaneous testimony and speeches, or the serious allegations of a remarkable pattern of deceptive conduct spelled out in the indictment. Moreover, none of the most recent reports indicates that defense experts fully evaluated Mr. Brockman's competency to stand trial using either conventional forensic interviewing or standardized competency to stand trial instruments.

I am now able to conclude with reasonable medical certainty that Mr. Brockman has sufficient present ability to consult with his attorneys with a reasonable degree of rational understanding and a rational as well as factual understanding of the proceedings against him.
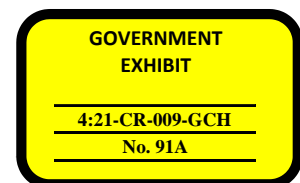
Respectfully,

Park Dietz, M.D., M.P.H., Ph.D.
Clinical Professor of Psychiatry and Biobehavioral Sciences
David Geffen School of Medicine at U.C.L.A.
Diplomate, American Board of Psychiatry & Neurology

EXHIBIT 91A

Agronin July Exam

Excerpt from Ex. 91, timestamp: 30:25 to 34:05

Message

From:	Stuart Yudofsky [stuart.yudofsky@gmail.com]
Sent:	5/4/2017 12:27:18 AM
To:	Bob Brockman [bob_brockman@reyrey.com]
Subject:	Anosmia

Good evening, Bob:

There are dozens of potential causes of anosmia--as there are for memory loss.

If you are comfortable doing so, I suggest that you begin by meeting with me. That will help me assess the severity and significance of your memory symptoms, as well as utilize this information to direct you to the best professional-- discipline, person, and/or system--for assessment and, as required, care.

As you know, Bob, I am a neuropsychiatrist and have extensive hands-on experience with memory and cognitive changes.

I am available in the mornings from 9 AM until noon over the next several days, and can be available at any time over the weekend.

Please let me know how you would like to proceed.

Warmest regards,

Stuart

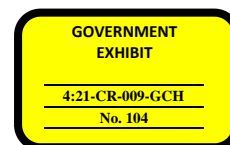On May 3, 2017, at 4:27 PM, Bob Brockman <bob_brockman@reyrey.com> wrote:

> Stuart,
>
> Robert and Dorothy are after me to consult with the right doctor regarding my loss of my sense of smell.
>
> They are afraid that it is an early sign of alzheimer's or dementia.
>
> I am feeling good but am having increasing memory problems.
>
> Is there a doctor that you can recommend?
>
> Bob

Ex. 104 p. 1 of 1

Message
_____

From: Bob Brockman [bob_brockman@reyrey.com]
Sent: 10/9/2019 3:12:37 AM
To: 'Don Passmore' [passmore@houston-cpa.com]
CC: Dorothy Brockman (Dorothy_Brockman@reyrey.com) [Dorothy_Brockman@reyrey.com]
Subject: RE: 2018 Tax Returns due 10/15/19 for Bob Brockman

Don

MAR was a first place prize winner in the Jones School Business Plan Competition. The prize is put up every year by the GOOSE group. Goose is the Grand Old Order of Successful Entrapreneurs. (sp)

The prize was $100K put up at $10K each -- I believe that there were 10 members.

The prospective investment was so appealing, that the GOOSE Group at the time it was tripled to $300K with each member putting up $30K instead of $10K.

Therefore I believe that my initial basis was $30K.

There were then additional capital calls that increased the basis in MAR.

My son Robert is working on digging out the details on these additional contributions to the basis of the investment -- as well as the returns to date. I am dove shooting in Argentina -- be back this coming Thursday evening.

Love,

Dad

PS: Robert -- this issue needs immediate attention as we are running out of time.

From: Don Passmore [mailto:passmore@houston-cpa.com]
Sent: Monday, October 07, 2019 12:15 PM
To: 'Bob Brockman'
Subject: 2018 Tax Returns due 10/15/19 for Bob Brockman

Bob
Was this a 2017 sale? Did you received proceeds in 2017 ($1,104,180)? I do not see anything in the 2017 tax information.
What was your cost in the investment $300,000? 2.6M?.
Please call me to discuss.
Thanks
Don

Don L. Passmore, CPA
Passmore & Associates, LLC
713-935-0300
713-935-0305 Fax

CONFIDENTIALITY STATEMENT:
This electronic message transmission contains information from Passmore & Associates, LLC and is confidential or privileged. The information is intended only for the use of the person(s) named above. If you are not the intended recipient, any use, disclosure, copying, distribution or any other action based on the contents of this email and any

Ex. 115 p. 1 of 4

attachments is strictly prohibited. If you received this electronic transmission in error, this message and any attachments should be deleted immediately and please notify us by telephone at 713-935-0300.

From: Robert Theron Brockman II [mailto:robert@firehead.org]
Sent: Monday, October 07, 2019 10:51 AM
To: Don Passmore
Cc: Daddy
Subject: 2018 Tax Returns due 10/15/19 for Bob Brockman

Ok, I think we've finally dug up enough documents to be able to answer these questions.

The original investment in MAR was a convertible note for 300k USD out of a total of 2.6M USD. It was sold as Adhesys Medical GmbH, a German LLC.

The holding has been 100% disposed of, but the payout schedule from the sale (including escrow, etc.) extends for some time. Estimated values are as follows, hopefully these at least roughly line up with what you've been seeing:

Closing: $1,104,180
Second: $150,816
Third: $75,408
Fourth: $377,039
Fifth: $301,631
Milestone 1 Escrow: $75,408
Milestone 2 Escrow: $75,408
Milestone 3 Escrow: $75,408

Attached is the pre-final draft of the sale and purchase agreement and an OCR scan of the original note text (best we can do for now.) Let me know if this is enough information.

Thanks,

Robert Brockman II

From: Don Passmore [mailto:passmore@houston-cpa.com]
Sent: Friday, September 27, 2019 4:58 PM
To: Laura Douglass; 'Bob Brockman'
Subject: 2018 Tax Returns due 10/15/19

Bob/Laura

We are completing the 2018 U S Individual Income Tax Return (Form 1040) and State returns due 10/15/19 for Bob and Dorothy and need assistance with the items:

For Bob's return

1.   Regarding the $300,000 investment in Medical Adhesive Revolution Inc.

a.   Was the investment for shares of stock in a "small business" corporation (total capital under 1MM).

b.   Did he dispose of 100% of his holding.

c.   If not a corporation what type of entity did he hold an interest.

d.   What underlying documents supported the investment.

Ex. 115 p. 2 of 4

2)   Number of days worked in OH in order to determine the OH portion of Wages; for extension assumed same as last year (3, 2 night trips to Dayton)

For Dorothy's return

1.   2018 Schedule K-1 Form 1065 from Versant Venture Capital V LP (Draft K-1 was provided and included in estimate)

2.   2018 Schedule K-1 Form 1065 from Versant Venture V AIV LP

Please call if you have any questions.

Thanks

Don

From: Laura Douglass [mailto:Laura_Douglass@reyrey.com]
Sent: Tuesday, May 07, 2019 3:41 PM
To: Don Passmore
Subject: RE: 2018 Tax Estimate and Extensions due 4/15/19

Don,

After looking through MANY files, I was able to find more info on #5 below.

A check for $300,000 was written to Medical Adhesive Revolution Inc. which is a part of Adhesys Medical Grunenthal. It was written and sent to them on 5/19/15.

That is the only payment I have found regarding this entity.

Laura
From: Don Passmore [mailto:passmore@houston-cpa.com]
Sent: Tuesday, April 09, 2019 9:41 AM
To: Bob Brockman <bob_brockman@reyrey.com>
Cc: Laura Douglass <Laura_Douglass@reyrey.com>
Subject: 2018 Tax Estimate and Extensions due 4/15/19

Bob
We have prepared 2018 tax estimates for both you and Dorothy based on the information provided. Based on the estimates both you and Dorothy are overpaid on your federal returns*.   The overpayments, along with expected withholding from 2019 wages should be enough to cover the 1st quarter 4/15/19 2018 ES tax payments based on the prior year tax. Therefore no payments are needed for the federal extensions and no federal ES payments are due 4/15/19. We will re-assess your 2018 estimate before the 6/15/19 ES payment.

*assumes the scheduled 1/15/19 Federal ES Payments were made timely of 210,687 for Bob and 89,827 for Dorothy

Based on the estimates both you and Dorothy are overpaid on your Ohio returns.   The overpayments should be enough to cover the 2019 ES tax payments based on the prior year tax. Therefore no Ohio extension payments are due 4/15/19. We will re-assess your 2019 Ohio estimate when we complete the 2018 return.

Based on the prior year the following is a list of information not available and not considered for the estimates:
Regarding Bob & Dorothy
1)   Current QuickBooks backup (QBB file) (upload if possible)
2)   OH portion of Wages; for extension assumed same as last year (3 2 night trips to Dayton)

Regarding Bob
1)   2018 Schedule K-1 Form 1065 from Hardwicke Properties, LLC
2)   2018 Schedule K-1 Form 1065 from Briteseed, LLC
3)   2018 Schedule K-1 Form 1065 from Nehemiah Ventures, LLC
4)   2018 Schedule K-1 Form 1065 from Nehemiah Holdings, LLC
5)   Date acquired and Cost Basis of Adhesys Medical Grunenthal P sold for 152,068 through Acquiom Financial, LLC
(included in estimate with no cost basis; capital loss carry over to 2018 offsets gain)
6)   Confirm Federal ES Payments were made timely of 210,687 1/15/19

Regarding Dorothy
1)   2018 Form 1099-DIV from Lord Abbett Affiliated A
2)   2018 Tax Reporting Statement from Fidelity account xx3715
3)   2018 Schedule K-1 Form 1065 from AP Energy Infrastructure Fund
4)   2018 Schedule K-1 Form 1065 from Versant Venture Capital V LP (Draft K-1 was provided and included in estimate)
5)   2018 Schedule K-1 Form 1065 from Versant Venture V AIV LP
6)   Confirm Federal ES Payments were made timely of 89,827 1/15/19

Also not considered for the estimates are any new flow-through entities for 2018 and any sales of investments outside of your brokerage accounts.

You do not need to sign either the federal or state extensions.

Please call me to discuss details of the estimate and potential penalties and interest on unpaid or under paid taxes.

Thanks
Don

Don L. Passmore, CPA
Passmore & Associates, LLC
713-935-0300
713-935-0305 Fax

CONFIDENTIALITY STATEMENT:
This electronic message transmission contains information from Passmore & Associates, LLC and is confidential or privileged. The information is intended only for the use of the person(s) named above. If you are not the intended recipient, any use, disclosure, copying, distribution or any other action based on the contents of this email and any attachments is strictly prohibited. If you received this electronic transmission in error, this message and any attachments should be deleted immediately and please notify us by telephone at 713-935-0300.

Ex. 115 p. 4 of 4
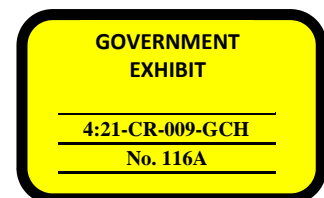
EXHIBIT 116

Brockman Dec. 2018 Reynolds & Reynolds Speech
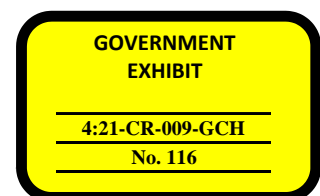
EXHIBIT 116A

Brockman Dec. 2018 Reynolds & Reynolds Speech
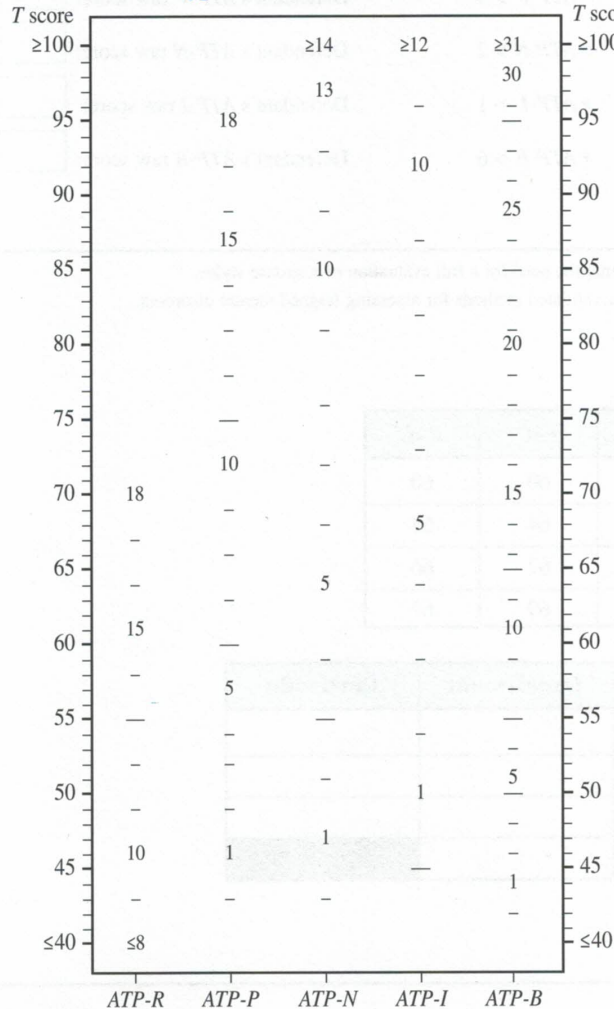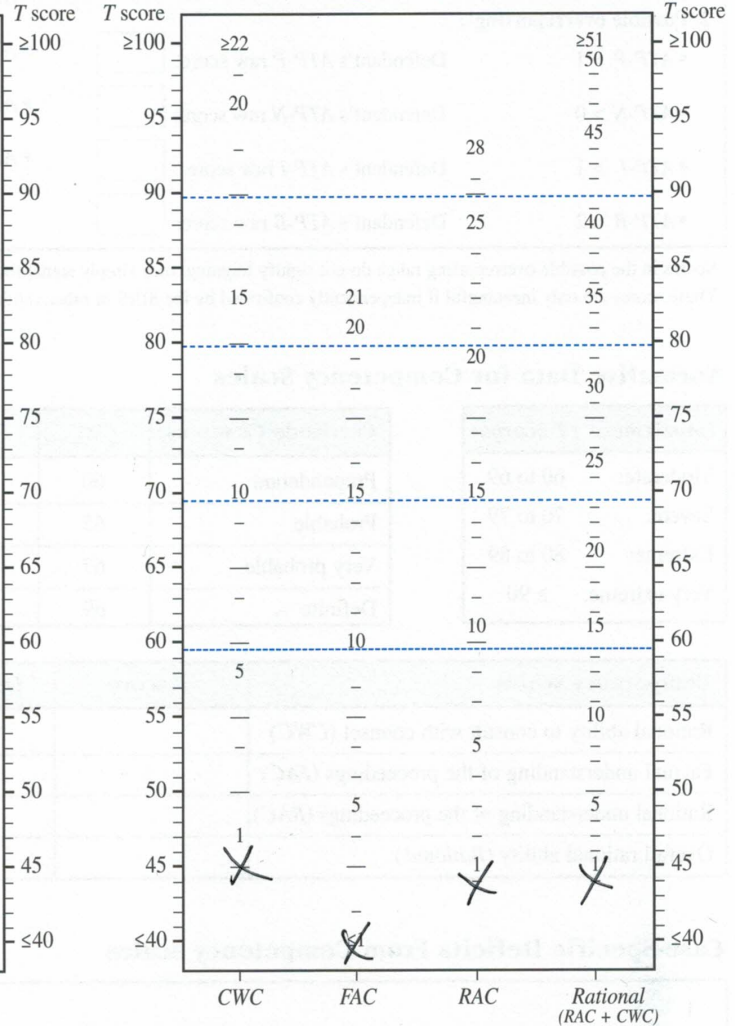
Excerpt from Ex. 116, timestamp: 24:20 to 26:11

ECST-R Profile Form

Name: R.T. Brockman   Age: 80   Testing date: 10/26/21 (Mo. Day Yr.)

Examiner: R.L. Denney   Place of evaluation: Jones Day Houston

Atypical Presentation scales

	ATP-R	ATP-P	ATP-N	ATP-I	ATP-B
T score					
Raw score					
%ile					

Competency scales

	CWC	FAC	RAC	Rational (RAC + CWC)
T score	45	40	44	44
Raw score	0	1	0	0
%ile				

WMT
Click Here for Summary

Criterion A
Details

IR, DR & CNS
ALL > 82.5%

Pass Easy Subtests

Free Recall ≥ 40 → **Normal Range Memory (But Compare With Relevant Comparison Groups)**

Free Recall < 40 → **Weak Memory. Interpret Relative To Appropriate Comparison Groups**

No

IR ≤ 82.5%
OR
DR ≤ 82.5%
OR
CNS ≤ 82.5%

Is IR, DR or CNS worse than chance?
Details

Yes

Order Violations
Details

Poor Effort Profile → **Memory Subtests Probably Invalid**

< 30

Fail Easy Subtests → **Criterion B (Easy - Hard)**
Details

≥ 30

Possible Genuine Memory Impairment → **Memory Subtests Could Be Valid (Apply Slick et al criterion D)**
CLICK HERE
→ **Interpret Relative To Appropriate Comparison Groups**

Legend:
A: (IR <= 82.5%) OR (DR <= 82.5%) OR (CNS <= 82.5%) ?
Met

B : (Easy - Hard Difference) < 30 ?
[(IR + DR + CNS) / 3] - [(MC + PA + FR) / 3] = **Not Met**

TestDate	IR	DR	CNS	MC	PA	FR	LDFR
5/19/2021	52.5	47.5	65	15	15	0	

Green's WMT (AI)

Legend:
- Patient: Brockman, Robert. Test administered: 5/19/2021
- ELDERLY WITH EXIT SCORE 0 to 9 (Dr. Brockhaus, N=21, mean age=67)
- ELDERLY WITH EXIT SCORE 10 to 14 (Dr. Brockhaus, N=51, mean age=70)
- DEMENTIA WITH EXIT SCORE 15 to 19 (Dr. Brockhaus, N=45, mean age=73)
- DEMENTIA WITH EXIT SCORE 20 to 24 (Dr. Brockhaus, N=50, mean age=76)
- DEMENTIA WITH EXIT SCORE 25+ (Dr. Brockhaus, N=31, mean age=76)
- Abnormal Performance Area

Green's WMT (AI)

NV-MSVT
Click Here for Summary

```
                          A1          > 90%        A2        > 88%    Pass Easy              Free Recall          Normal Range Memory
                       Details                   Details                Subtests                ≥ 70           (But Compare With Relevant
                                                                                                                 Comparison Groups)
              No
                             ≤ 90%         ≤ 88%                                             Free Recall               Weak Memory.
                                                                                               < 70               Interpret Relative To
       Is IR, DR, CNS,                                                                                          Appropriate Comparison
       DRA or DRV                                                                                                       Groups
       worse than
         chance?                                                      Fail All 3 B          Poor Effort Probable:
                                                                       Criteria          This profile has close to zero
       Details                                                                           false positives in very severe           Order
                                                                                          impairment from dementia               Violations
              Yes                              Fail A                                                                           Details
                                               Criteria               Fail Any 2          Poor Effort Probable:
                                                                      B Criteria       This profile has relatively few
                                            Criterion B1 Details                         false positives in dementia
       Poor Effort                          Criterion B2 Details
        Profile                                                       Fail 1 B             Data Could be Valid
                                            Criterion B3 Details      Criterion           but apply Slick et al          CLICK HERE
                                                                                            criterion D
```

Legend:
A1: (IR + DR + CNS + DRA + DRV + PA)/6 <= 90 ? Met
A2 : (DR + CNS + DRA + DRV)/4 <= 88 ? Met
B1 : PA - (DR + CNS + DRA + DRV)/4 > -11 ? Not Met
B2 : (IR + DR + CNS)/3 - (PA + FR)/2 < 20 ? Not Met
B3 : SD(IR, DR, CNS, DRA, DRV) >= 12 ? Not Met

Fail 0 B Criteria → **Data Could be Valid but apply Slick et al criterion D**   CLICK HERE

TestDate	IR	DR	CNS	DRA	DRV	PA	FR	Effort?
5/19/2021 9:11:20 AM	90	80	80	90	100	20	0	Yes

Green's NV-MSVT (AI)

Legend:
- Patient: Brockman, Robert. Test administered: 5/19/2021 9:11:20 AM
- 48 Sophisticated Dementia Simulators Armistead-Jehle & Denney AN-Adult 2015
- 33 MMC Clinical Dementia Cases
- 49 dementia patients, mean age=78 (SD=7) Dr. Cohen (N=49)
- Mixed dementia group who passed NV-MSVT criterion A (N=56, Dr. Cohen)
- Mixed dementia group who failed NV-MSVT criterion A (N=55, Dr. Cohen)
- Abnormal Performance Area

Determination of the smoking gun of intent: significance testing of forced choice results in social security claimants

Laurence M. Binder & Michael D. Chafetz

Published online: 15 Jun 2017.

Submit your article to this journal

Article views: 206

View related articles

View Crossmark data

GOVERNMENT EXHIBIT

4:21-CR-009-GCH
No. 122

GOVERNMENT EXHIBIT
122

Routledge
Taylor & Francis Group

Check for updates

CE Determination of the smoking gun of intent: significance testing of forced choice results in social security claimants

Laurence M. Binder[a‡] and Michael D. Chafetz[b‡]

[a]Independent Practice, Beaverton, OR, USA; [b]Independent Practice, New Orleans, LA, USA

ABSTRACT

Objective: Significantly below-chance findings on forced choice tests have been described as revealing "the smoking gun of intent" that proved malingering. The issues of probability levels, one-tailed vs. two-tailed tests, and the combining of PVT scores on significantly below-chance findings were addressed in a previous study, with a recommendation of a probability level of .20 to test the significance of below-chance results. The purpose of the present study was to determine the rate of below-chance findings in a Social Security Disability claimant sample using the previous recommendations. **Method**: We compared the frequency of below-chance results on forced choice performance validity tests (PVTs) at two levels of significance, .05 and .20, and when using significance testing on individual subtests of the PVTs compared with total scores in claimants for Social Security Disability in order to determine the rate of the expected increase. **Results**: The frequency of significant results increased with the higher level of significance for each subtest of the PVT and when combining individual test sections to increase the number of test items, with up to 20% of claimants showing significantly below-chance results at the higher p-value. **Conclusions**: These findings are discussed in light of Social Security Administration policy, showing an impact on policy issues concerning child abuse and neglect, and the importance of using these techniques in evaluations for Social Security Disability.

Introduction

In neuropsychological and medical evaluations, examinees may malinger by intentionally underperforming on measures of ability or by exaggerating symptoms (Heilbronner et al., 2009). Malingering is defined as the intentional production or exaggeration of symptoms or abnormal behavior, or the manipulation of test performance, in order to deceive the examiner into incorrectly concluding that the claimed dysfunction is greater than what actually exists, for the purpose of obtaining an external incentive (Slick, Sherman, & Iverson, 1999; Bianchini, Greve, & Glynn, 2005; Heilbronner et al., 2009; Chafetz et al., 2015). For the incentive of Social Security Disability (SSD) benefits, claimants who are malingering may

---

CONTACT Michael D. Chafetz ✉ brain.resource@yahoo.com
‡Authorship is alphabetical.

exaggerate their functional limitations in order to meet SSD disability requirements ((Chafetz et al., 2015). Performance validity tests (PVTs) are useful for assessment of the validity of cognitive testing (Armistead-Jehle & Denney, 2015; Bianchini et al., 2005; Larrabee, 2012). However, for several years, the Social Security Administration (SSA) has actively discouraged the use of validity testing in its consultative examinations (Chafetz et al., 2015) for reasons that are outside the scope of this paper (see Chafetz, 2010).

PVTs that employ established cut-offs are designed to assess the validity of performance in neuropsychological evaluations (Larrabee, 2012). Many PVTs utilize a forced choice format in which the examinee must choose between targets and foils. Apart from the use of established cut-offs, any forced choice test score falling below chance can also be tested for statistical significance. Testing determines if the score is sufficiently low that it is unlikely it could have been generated by unlucky guessing of the incorrect answers by truly impaired individuals. If the result is statistically significant, then it is highly probable that the examinee intentionally generated incorrect answers (Binder, 1990; Pankratz & Erickson, 1990; Slick et al., 1999).

A PVT score that is significantly below chance provides the strongest psychometric evidence of deliberate underperformance that is consistent with a diagnosis of malingering because it provides evidence that the examinee intentionally gave wrong answers. The alternative explanation that this score was obtained by consistently unlucky guessing of the answers by impaired individuals is less likely because the cutting scores obtained on samples composed of impaired individuals with well-documented brain injuries or disease are well above chance. For example, the below-chance score might be compared to the much higher scores obtained in a sample with well-documented moderate to severe traumatic brain injuries (Tombaugh, 1996).

The original Slick et al. (1999) criteria for the determination of malingering, as well as the revised criteria (Slick & Sherman, 2013), indicated that a PVT score significantly below chance at the .05 level provides 'definite' evidence of malingering in the context of external incentives and symptoms and behaviors that cannot be fully explained by neurological, psychiatric, or developmental factors (D Criteria). Pankratz and Erickson (1990) described a significantly below-chance result on a forced choice PVT as 'the smoking gun of intent' because a higher (chance-level) score would be obtained if the claimant was blind to the choices and guessed the answer for each item. Guessing is likely to generate a chance-level performance. Intentionally providing a sufficient number of wrong answers leads to a score that is significantly below chance.

Chafetz (2008) provided base-rate evidence of the frequency of significantly below-chance PVT failure in SSD claimants for Supplemental Security Income (SSI; children and adults) and Social Security Disability Insurance (SSDI; adults) income. This study showed that 12.3–13.6% of adult claimants and 10% of child claimants had PVT scores that were significantly below chance, defined with a binomial probability level of $p < .05$. As these figures provide unequivocal evidence of a large number of adult and child claimants who are deliberately misrepresenting their claimed abilities in a public disability program, they potentially have policy implications for the SSA system.

Given the medicolegal importance of statistically significant below-chance results on PVTs, it is important to optimize the method of testing of significance. Binder, Larrabee, and Millis (2014) addressed the issues of probability levels, one-tailed vs. two-tailed tests, and combination of PVT scores. The probability level of .05 used for testing hypotheses about

134   L. M. BINDER AND M. D. CHAFETZ

issues such as the effects of pathogens and treatments or the associations between historical factors and dependent variables, using two or more independent samples from different populations, had been arbitrarily applied to the completely different statistical issue of determining if a forced choice test result was significantly worse than chance. As Binder et al. (2014) pointed out, the significance level for forced choice testing was appropriated from research on groups without receiving a logical or statistical test of its accuracy or appropriateness. The widespread acceptance of a probability level of .05 for comparison of groups was not a sufficient reason to use it for analysis of the results of a forced choice test in a single case. To determine an empirically based probability level for testing significance, they examined data from the Warrington Recognition Memory Test-Words (WRMT-W) from a sample of 127 adult participants with traumatic brain injuries (TBIs) ranging from mild to severe, none of whom had financial incentives for their test performance. None scored worse than 23/50 correct (46%). The cut-off score from this sample of 22/50 yielded a one-tailed p level of .24 when the binomial was calculated, rather than estimated from the z approximation. This same cut-off score was also supported on the WRMT-W in another sample of participants with mixed etiologies cited by Binder et al. (Kim et al., 2010). Based on these data, Binder et al. recommended a probability level of .20 for testing the hypothesis that a below-chance score was significant.

Binder et al. recommended one-tailed significance testing for below-chance scores to test the hypothesis that the results were significantly below chance. Their review of the literature showed no consensus about the issue of one- or two-tailed testing. They reasoned that the hypothesis of below-chance performance was uni-directional and one-tailed; there was no need for significance testing when scores were above chance, as low scores that were 50% correct or better were compared with empirically based cut-off scores.

Combining PVT scores also was recommended in order to increase the number of items and thus the statistical power of significance testing with the binomial theorem (Binder et al., 2014). Scores from tests divided into more than one subtest could be added together. Consistent with the recommendation of the authors of the Victoria Symptom Validity Test (VSVT; Slick, Hopp, Strauss, & Thompson, 2005) and of the Portland Digit Recognition Test (PDRT; Binder, 2007), significance could be tested on either section of the tests (easy or hard items) or on the total score. As long as they are independent, subtest scores could be combined (Binder et al., 2014) to increase statistical power. For example, on the Test of Memory Malingering (TOMM; Tombaugh, 1996), with three sections, each containing 50 items, for 21 correct of 50 items (42%), $p = .161$. If the same percentage correct (42%) was obtained on the total score for all three sections of the test (63 correct of 150 items), $p = .03$.

Subtests that are not independent cannot be combined. For example, the Word Memory Test (WMT; Green & Astner, 1995) and the Medical Symptom Validity Test (MSVT; Green, 2004) have two forced choice subtests: Immediate Recognition (IR) and Delayed Recognition (DR). Scores on IR and DR are independent and can be combined. However, the Consistency score for both tests is dependent on IR and DR subtest behavior, and thus cannot be combined with these other scores to test the significance of below-chance performance.

As reviewed in Binder et al. (2014), most reports of the frequency of significantly below-chance findings in the literature were based on p levels of .05. Both one-tailed and the more conservative two-tailed tests of significance were reported. While most of these reports are about single subtests, some investigators (Binder, 2007; Green, 2008; Slick et al., 2005) have

taken the opportunity to combine test scores when reporting the frequency of significantly below-chance scores.

While Chafetz (2008) already reported on the frequency of significantly below-chance ($p < .05$) findings using the TOMM and MSVT in SSD samples of adults and children, this prior study utilized only this conservative probability level, analyzing only single (and not combined) subtests. The present study was designed to determine the frequency of significantly below-chance findings using more recent, empirically derived probability levels while also examining combined subtest scores. Obviously, a higher p level of .20 as recommended by Binder et al. (2014) will produce a higher frequency of significantly below-chance findings, but it was of interest to determine just how much higher. As these findings are potentially of interest for SSA policy concerning misrepresentation of a claimant's disabling problems, they are discussed in relation to SSA rules. Policy concerning child disability was also of concern, as Chafetz and Dufrene (2014) have implicated below-chance PVT findings in child abuse and neglect.

Methods

Participants

As previously described in Chafetz, Abrahams, and Kohlmaier (2007) and Chafetz (2008), archived records from consecutive Disability Determinations Services (DDS) referrals to the practice of Michael Chafetz for the psychological consultative examination (PCE), most alleging low cognitive functioning, were used. When the claimants signed the HIPAA notification, they were asked for permission to use the scores from their examination for research and assured that their identity would be protected. Permission was granted if the claimant initialed next to the research notification. Only one claimant declined, and her scores were not used. The first author presented preliminary findings of this research at a statewide Louisiana DDS meeting (April 28, 2003), and the local DDS Medical Liaison Officer provided written acknowledgment about this presentation and the research involved.

Table 1 shows the demographic tabulations of the records. As indicated (Chafetz, 2008), participants in the TOMM study were 232 consecutive DDS referrals for the PCE. Of the 221 referrals who were administered the TOMM, 129 were WAIS-III-age, and 92 were WISC-III-age, and the two data-sets were separated into samples of adults and children. When DDS shifted to the use of the WISC-IV, the MSVT was then used for validity testing. In the MSVT study, there were 55 WAIS-III-age participants and 27 WISC-IV-age participants, which were again separated into samples of adults and children. As the present study had different objectives

Table 1. Sample characteristics for adults and children on the Test of Memory Malingering and Medical Symptom Validity Test (adults only) studies.

	Test of Memory Malingering		Medical Symptom Validity Test
	Adult	Child	Adult
N	129	92	55
Age[†]	28.9 (10.9)	10.7 (2.7)	29.6 (12.1)
Education[†]	9.6 (2.0)	3.6 (2.4)	9.3 (1.9)
Gender	71 M, 58F	64 M, 28F	33 M, 22F
Ethnicity	103B, 24 W, 2O	81B, 11 W	48B, 7 W

Notes: B = Black/African-American; W = White; O = Other; M = Male; F = Female.
[†]M (SD).

than in Chafetz (2008), there were differences in the number of records involved. For example, the Chafetz (2008) study had more restrictive inclusion rules for the frequencies in each condition, requiring each case to have all PVT and IQ data. Except for the Adult MSVT table, the tables that show the rate of significantly below-chance findings in the base-rate study (Chafetz, 2008) have fewer cases than in the present study. The Child MSVT data were not included in the current study due to low sample size.

Performance validity tests

TOMM (Tombaugh, 1996): In the original SSD study (Chafetz et al., 2007), two trials of the TOMM were used in the formal testing of performance validity. The TOMM is not described here because it is well known to the readers. The Retention trial was not used. MSVT: The MSVT is a forced choice test that contains a list of 10 pairs of words presented twice to each claimant on a computer screen. Claimants are then asked to choose the correct target words from pairs consisting of a target and a foil. Because most claimants – adults and children – complained of reading difficulties, the examiner read all the words and choices on the screen, thereby effecting a combined computer and oral administration (Chafetz, 2008). This accommodation of impaired claimants, removing a possible source of error from reading or executive dysfunction and leaving only the claimant's (forced) choices, was further discussed in Chafetz (2015). We know of no research examining the effect of this modification of the standard administration, though we note that the origin of the WMT changed from oral to computer administration without change of cut-offs, and the equivalence of the oral and computer administrations has been demonstrated (Hoskins, Binder, Chaytor, Williamson, & Drane, 2010). However, any hypothetical change of scores from this hybrid method would likely be in the positive direction (e.g. eliminating impulsive responding errors), which would render our results in this study less likely to be due to inadvertent responses in truly impaired individuals. Moreover, Dr. Paul Green, the author of the MSVT, suggested in a personal communication (May 11, 2017) that this hybrid method would assist the very few with severe reading difficulties, thereby minimizing failures.

Binomial probability significance levels

Statistical significance was tested with the binomial theorem, as calculated online at the Vassarstats website http://vassarstats.net/textbook/ch5apx.html, which provides an exact binomial probability calculator. The exact binomial probability (P_b) is obtained using various applications of the formula:

$$P_b(k \text{ out of } N) = (N!/k!(N - k)!) \, (p^k)(q^{N-k})$$

In this formula: N = number of opportunities for an event to occur (e.g. choosing the target word), k = number of times that the event occurs or is stipulated to occur; in this case, the number of correct target choices, p = probability the event will occur on any particular occasion (e.g. .5 for 2 forced choice), q = probability the event will not occur on any particular occasion (e.g. .5 for 2 forced choice).

We used the P_b value for a one-tailed test, applying the formula to all values of k equal to or smaller than the stipulated k amount, which is the probability of getting k or fewer

events. For the MSVT and the TOMM (or any other PVT with two forced choice alternatives), Pb refers to the probability of obtaining such a result by chance. For example, for 20 (or fewer) correct of 50 items on the TOMM, k = 20, N = 50, the one-tailed exact binomial probability (P_b) is .1013.

Results

Adult sample

In the adult MSVT sample, 7 (12.7%) individuals obtained scores on Immediate Recognition (IR) or Delayed Recognition (DR) that were worse than chance at the .05 level, compared with 9 (16.4%) individuals at the .20 level (see Table 2). When MSVT IR and DR scores were summed, 6 (10.9%) individuals performed significantly less than chance at .05, compared with 8 (14.5%) at .20. Considering the typical cut-off on the MSVT (<90%), 33 individuals (60%) performed worse than the cut-off score on either IR or DR.

For either trial on the TOMM, 18 (14%) individuals scored worse than chance at the .05 level, compared with 26 (20.2%) at the .20 level. When scores on both trials were summed, the number of participants with significantly below-chance scores increased from 19 (14.7%) at the .05 level to 22 (17.1%) at the .20 level. For the conventional level of failure on the TOMM (Trial 2 < 90%), 72 individuals (55.8%) scored below the cut-off (see Table 3). The reader will notice from the tables that the number of individuals in the summed trials with significantly below-chance results was the same as the highest MSVT trial alone and between the two TOMM trials. While the statistical power increases with summed trials, this finding is an expected result when one of the trials has several fewer claimants obtaining below-chance results.

Child sample

For either of the TOMM trials, 11 (12%) children scored worse than chance at the .05 level, compared with 12 (13%) children at the .20 level. Summing the scores on Trials 1 and 2 yielded 8 (8.7%) children at the .05 level and 11 (12%) children at the .20 level. For the traditional failure cut-off (<90%) on Trial 2 of the TOMM (used in Chafetz et al., 2007), 26 (28.3%) children scored below cut-off (see Table 4).

Table 2. Frequencies of significantly below-chance scores at different binomial probability levels: adults MSVT.

Cut score (≤X)	Percent correct (20 trials)	Binomial Probability	IR% (#)	DR% (#)	IR or DR% (#)
MSVT IR or DR (20 trials)					
5	25	.0207	9.1 (5)	7.3 (4)	12.7 (7)
7	35	.1316	14.5 (8)	10.9 (6)	16.4 (9)
17	85	.9998	52.7 (29)	58.2 (32)	60 (33)
Cut score (≤X)	Percent (40 trials)	Binomial Probability	IR + DR% (#)		
MSVT IR plus DR (40 trials)					
14	35	.0403	10.9 (6)		
16	40	.1341	14.5 (8)		

Notes: MSVT = Medical Symptom Validity Test; TOMM = Test of Memory Malingering; IR = Immediate Recognition; DR = Delayed Recognition; T1 = Trial 1; T2 = Trial 2. In these tables, the one-tailed binomial probabilities are given for all successes (# items correct) up through the cut score. For example, if the cut score is ≤7, the binomial probability for 7 or fewer successes is given. Traditional failure cut score is less than 90% (for TOMM just T2). Adults (MSVT: n = 55; TOMM: n = 129).

138   L. M. BINDER AND M. D. CHAFETZ

Table 3. Frequencies of significantly below-chance scores at different binomial probability levels: adults TOMM.

Cut Score (≤X)	Percent (50 trials)	Binomial Probability	T1% (#)	T2% (#)	T1 or T2% (#)
TOMM T1 or T2 (50 Trials)					
18	36	.0325	6.2 (8)	13.2 (17)	14.0 (18)
21	42	.1611	11.6 (15)	19.4 (25)	20.2 (26)
44	88	.9999	–	55.8 (72)	–
Cut Score (≤X)	Percent (100 trials)	Binomial Probability	T1 + T2 % (#)		
TOMM T1 plus T2 (100 trials)					
41	41	.0443	14.7 (19)		
45	45	.1841	17.1 (22)		

Notes: MSVT = Medical Symptom Validity Test; TOMM = Test of Memory Malingering; IR = Immediate Recognition; DR = Delayed Recognition; T1 = Trial 1; T2 = Trial 2. In these tables, the one-tailed binomial probabilities are given for all successes (# items correct) up through the cut score. For example, if the cut score is ≤ 7, the binomial probability for 7 or fewer successes is given. Traditional failure cut score is less than 90% (for TOMM just T2). Adults (MSVT: $n = 55$; TOMM: $n = 129$).

Table 4. Frequencies of significantly below-chance scores at different binomial probability levels: children TOMM.

Cut Score (≤X)	Percent (50 trials)	Binomial Probability	T1% (#)	T2% (#)	T1 or T2% (#)
TOMM T1 or T2 (50 trials)					
18	36	.0325	6.5 (6)	12 (11)	12 (11)
21	42	.1611	8.7 (8)	12 (11)	13 (12)
44	88	.9999	–	28.3 (26)	–
Cut score (≤X)	Percent (100 trials)	Binomial Probability	T1 + T2% (#)		
TOMM T1 plus T2 (100 trials)					
41	41	.0443	8.7 (8)		
45	45	.1841	12 (11)		

Notes: TOMM = Test of Memory Malingering; IR = Immediate Recognition; DR = Delayed Recognition; T1 = Trial 1; T2 = Trial 2. In these tables, the one-tailed binomial probabilities are given for all successes (# items correct) up through the cut score. For example, if the cut score is ≤ 7, the binomial probability for 7 or fewer successes is given. Traditional failure cut score is less than 90% (for TOMM just T2). Children (TOMM: $n = 92$).

Discussion

The use of the less conservative probability level for significance testing of .20 led to a small increase in the frequency of significant results, compared with a probability of .05. In the adult MSVT sample, for either recognition trial, the increase was from 12.7 to 16.4%. Using the combined MSVT recognition trials, the increase was from 10.9 to 14.5%. In the TOMM adult study, using either trial, the increase was from 14 to 20.2%. With summed trials, the increase was from 14.7 to 17.1%. In the child TOMM study, using either trial, the increase was from 12 to 13.0%. With summed trials, the increase was from 8.7 to 12%. In comparison, on the VSVT, in a sample not known to have external incentives for poor performance, 0.3% were below chance at the .05 level and 2.3% at the .20 level (Loring, Larrabee, Lee, & Meador, 2007). The current results also are consistent with a prior estimate (Binder et al., 2014) that was based on data from the PDRT (Greve & Bianchini, 2006).

When testing the level of significance, we recommend the exact calculation of the binomial, using an online calculator such as VassarStats, rather than the z approximation that was used in the prior paper (Binder et al., 2014). In some instances, the approximation gives a non-precise estimate.

Both the original Slick criteria (Slick et al., 1999) and the revised criteria (Slick & Sherman, 2013) specified a .05 level of significance for definite malingering. Binder et al. (2014)

proposed broadening this criterion to .20, based on statistical principles, cutting scores, and data from two samples of brain injured persons, including one sample found in the literature (Kim et al., 2010). Of potential concern is the possibility that impaired individuals who guess on all items are more likely to obtain results significantly worse than chance with the higher p level of .20. Such performances, if they were to occur, should not be classified as deliberate provision of wrong answers. Our current data on the frequency of below-chance results in our sample do not directly address the validity of the higher .20 probability level.

However, the empirical literature does not support the concern that misclassification is likely with the higher probability level in clinical samples with no known external incentives for poor performance. We reviewed forced choice test data that could be extracted for individual participants from several papers and two test manuals (Binder, 2007; Binder & Kelly, 1996; Carone, 2008; Chafetz & Biondolillo, 2012; Greve & Bianchini, 2006; Greve, Bianchini, & Doane, 2006; Haber & Fichtenberg, 2006; Howe, Anderson, Kaufman, Sachs, & Loring, 2007; Kim et al., 2010; Loring et al., 2007; MacAllister, Nakhutina, Bender, Karantzoulis, & Carlson, 2009; Macciocchi, Seel, Alderson, & Godsall, 2006; Schroeder et al., 2012; Singhal, Green, Ashaye, Shankar, & Gill, 2009; Slick et al., 2003). We also contacted one of these authors of a pediatric study and obtained previously unpublished details about the results of individual cases in one published study cited here on the TOMM, additional unpublished data on 93 pediatric cases that followed the original study of the TOMM, and 25 unpublished pediatric cases with the MSVT (William MacAllister, personal communications, February 23 and March 1, 2017). In a published study of children with fetal alcohol syndrome (Larson et al., 2015), 6 of 120 cases performed below the usual cut-offs on the MSVT; we were unable to learn via personal communication how many cases performed below chance at the .20 level, and we did not include this study in our aggregate summary. Some of the samples in the studies reviewed above include children with low intelligence. In aggregate, including the unpublished cases, 9 of 1630 cases (0.6%) with no known external incentives for poor performance scored worse than chance ($p < .20$) on a forced choice test.

The proportion of individuals obtaining significantly worse than chance results at the .20 level in these published and unpublished studies is much less than the proportion of .20 that would be expected if everyone was guessing the answers to the PVTs. We analyzed data from the Loring et al. (2007) study because this was the study with the highest frequency of below-chance responding; 8 of the 346 people (2.3%) in the clinical sample performed worse than chance. The obtained percentage in the Loring et al. (2007) sample is significantly less than the 20% expected if every participant guessed on all items, $z = -8.24$, $p < .00001$. For all 1630 cases, $z = -19.62$. If below-chance results in these published and unpublished samples result from random guessing, then only a small proportion of the participants are guessing. It also is possible that some of the participants in these studies who were classified as lacking external incentives actually were misclassified and had external incentives for their scores.

We also reviewed some of the experimental studies that asked participants to simulate cognitive impairment (Armistead-Jehle & Denney, 2015; Binder & Willis, 1991; Frazier, Frazier, Busch, Kerwood, & Demaree, 2008; Green, Flaro, & Courtney, 2009; Powell, Gfeller, Hendricks, & Sharland, 2004; Singhal et al., 2009; Vickery et al., 2004). The mean scores on forced choice PVTs for groups simulating cognitive impairment in these seven studies including nine groups were above 50% correct, except for a group simulating reading disorder on the VSVT

140   L. M. BINDER AND M. D. CHAFETZ

hard items (Frazier et al., 2008) and a group simulating TBI on the PDRT hard items that averaged just under 50% correct (Binder & Willis, 1991).

In summary, experimental studies have demonstrated that most simulators instructed to feign brain dysfunction perform better than chance. The clinical studies have shown that only 0.6% of various samples with no known external incentives, many of them with obvious cognitive impairment, performed significantly worse than chance ($p < .20$). Therefore, significantly below-chance findings ($p < .20$) have great clinical significance.

We found that a high rate of SSD claimants performed worse than chance at the .20 level, with up to 20% of adult claimants and up to 13% of child claimants performing significantly worse than chance on a performance validity test. The performance validity tests, in isolation, do not necessarily indicate that the examinees were malingering. The diagnosis of malingering includes the criterion of an external incentive for poor performance, which also is met in all SSD claimants.

Our findings potentially have an impact on policy and practice within the SSA. Administrative Law Judge Michael Gilbert (Personal Communication, 2016) highlighted that SSA and the Office of the Inspector General (OIG) have developed the Cooperative Disability Investigations (CDI) program (https://oig.ssa.gov/cooperative-disability-investigations-cdi), which works with the state DDSs and other departments of SSA to prevent and investigate fraud in various programs, including disability. Since the CDI program was established in 1998, investigative efforts have resulted in a reported $3.2 billion in projected savings to SSA's disability programs. Judge Gilbert indicated that it is unlikely that many of these investigations have been about malingering per se. Most investigations concerned issues such as inconsistencies in the record, concealing work after getting benefits, college activities, other activities of daily living, or lying about other facts of the claim. The use of this neuropsychological performance validity technology could benefit the CDI investigations if the DDSs were to refer based upon consultative examination findings.

We emphasize that psychologists do not have the legal authority to equate any of their findings with a determination of fraud. The authors are merely providing education to the fact-finders (e.g. administrative law judges) within SSA for their own determinations or referrals to OIG. The mere evidence of significantly below-chance results, along with other information in the record, may demonstrate a pattern that is factually meaningful to the fact-finder charged with assessing myriad data points throughout the medical record (Michael Gilbert, Personal Communication, 2016).

It is important to determine the practical significance of the current findings, as potentially the relaxation of the below-chance criterion may have an impact on SSA's assessment of its evidentiary requirements. This is a difficult proposition. The publicly reported actuarial tables of adult disabled beneficiaries by diagnostic group (only mental disorders) show that in SSDI there were 2,727,043 Disabled Worker Beneficiaries and in SSI 2,819,246 Adult Recipients, ages 18–64 for 2015 (SSA, 2016a, 2016b). Compared to the 2011 numbers reported in Chafetz and Underhill (2013), the 2015 SSI report did not break out the numbers by diagnostic category. Therefore, the previously reported percentage of mental disorders (58.6% of adult SSI cases) was used. Under the current rules, these figures represent beneficiaries for whom no validity evidence was systematically obtained (Chafetz, 2010; Chafetz et al., 2015). However, not all beneficiaries were sent for PCEs from which the current below-chance data were obtained, and it is known within the system that not all consultants make their recommendations fully from the results of the PCEs.

Another way of examining this problem is to study the disposition of applications for SSI disability benefits in 2015 (Table V.C1; SSA, 2016b). Of the 1,426,419 SSI cases filed in 2015, with 1,167,472 total decisions in that year, 29% were provided allowances (with some occurring after appeals and reconsiderations), while 71% were denied for various reasons (including: the impairment was not expected or did not last 12 months; the impairment was not severe or did not cause severe functional limitations; the claimants were able to do their past work or another type of work; or there was insufficient evidence, failure to cooperate, did not want to continue the claim, or returned to work). Of these denials, 17.7% (206,463) were still on appeal for the next step in the appeals process. The knowledge that up to 20% of claimants sent for PCEs to garner evidence for the fact-finding in these cases were showing 'the smoking gun of intent' to misrepresent impairment would have been informative in any stage of the process, including the further appeals at the Administrative Law Judge level.

The high rate of significantly below-chance findings in the child sample is of special concern. Feigning at the direction or pressure by others in the context of a financial incentive has been labeled malingering-by-proxy (MBP) (Slick et al., 1999; Slick & Sherman, 2013). As Chafetz and Dufrene (2014) asserted from their reviewed cases and from prior examples (see Cassar, Hales, Longhurst, & Weiss, 1996), MBP can be a form of abuse and may be treated as such by protective agencies. Chafetz and Dufrene (2014) proposed that in cases of MBP, which are especially signaled by below-chance findings, that the cases be referred to the local Child Protection Agency for investigation as a matter of policy.

One of the limitations of the current study comes from the definition of impairment in the study of Binder et al. (2014) as ranging from mild to severe TBI, without knowing the percentages or impact of the TBI levels. While none of the TBI subjects, including severe TBI, obtained a below-chance finding beyond the $p < .20$ level, these TBI subjects are not likely comparable to the claimants with low intellectual functioning in the current SSD sample. However, this is not likely to have created any interpretive problem in the current study, as it is known that well-motivated claimants with low intellectual functioning (IQ: 60–75) do not typically fail the MSVT or other PVTs designed for low-IQ claimants, much less obtain below-chance scores (Chafetz & Biondolillo, 2012). Furthermore, in this study, we aggregated literature showing that below-chance PVT findings at a probability level of .20 had a frequency of only 0.6% in clinical samples.

As with any study, it is possible that the frequency of below-chance findings is sample dependent. The frequency of below-chance findings should be reported in future studies of different populations. Another limitation is the non-random placement of targets and foils in the MSVT. A more conservative approach to the testing of significance may be indicated for tests with topographic locations of targets and foils that are non-random.

Neuropsychologists are aware that malingering at any level cannot be defined by a test result alone, and that it takes adherence to established guidelines (e.g. Slick & Sherman, 2013) for appreciation of the context of the findings. Moreover, we caution that neuropsychologists should not liberalize the established cut-offs without carefully validated empirical justification, such as Greve and Bianchini's work on the TOMM (2006).

In summary, the recommended use of a probability level of .20 for determining if a forced choice test score is significantly below chance (Binder et al., 2014) led to a small but meaningful increase in the frequency of significant results in disability claimants. Significant results in the context of an external incentive likely signify deliberate efforts to mislead the examiner,

142   (⊕)   L. M. BINDER AND M. D. CHAFETZ

i.e. malingering. The frequency of significant results in this sample of applicants for Social Security benefits has implications for the administration of these benefits.

Disclosure statement

Laurence Binder is the author and publisher of the Portland Digit Recognition Test. Michael Chafetz authored *Intellectual Disability: Criminal and Civil Forensic Issues,* published by Oxford University Press, from which he receives royalties. The authors perform medicolegal evaluations.

References

Armistead-Jehle, P., & Denney, R. L. (2015). The detection of feigned impairment using the WMT, MSVT, and NV-MSVT. *Applied Neuropsychology Adult, 22,* 147–155.

Bianchini, K. J., Greve, K. W., & Glynn, G. (2005). On the diagnosis of malingered pain-related disability: Lessons from cognitive malingering research. *The Spine Journal, 5,* 404–417.

Binder, L. M. (1990). Malingering following minor head trauma. *Clinical Neuropsychologist, 4,* 25–36.

Binder, L. M. (2007). *Portland digit recognition test manual* (3rd ed.). Beaverton, OR: Author.

Binder, L. M., & Kelly, M. P. (1996). Portland digit recognition test performance by brain dysfunction patients without financial incentives. *Assessment, 3,* 403–409.

Binder, L. M., Larrabee, G. J., & Millis, S. R. (2014). Intent to fail: Significance testing of forced choice test results. *The Clinical Neuropsychologist, 28,* 1366–1375.

Binder, L. M., & Willis, S. C. (1991). Assessment of motivation after financially compensable minor head trauma. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3,* 175–181.

Carone, D. A. (2008). Children with moderate/severe brain damage/dysfunction outperform adults with mild-to-no brain damage on the medical symptom validity test. *Brain Injury, 22,* 960–971.

Cassar, J. R., Hales, E. S., Longhurst, J. G., & Weiss, G. S. (1996). Can disability benefits make children sicker [Letter to the editor]? *Journal of the American Academy of Child and Adolescent Psychiatry, 35,* 700–701.

Chafetz, M. D. (2008). Malingering on the social security disability consultative exam: Predictors and base rates. *The Clinical Neuropsychologist, 22,* 529–546.

Chafetz, M. D. (2010). Symptom validity issues in the psychological consultative examination for social security disability. *The Clinical Neuropsychologist, 24,* 1045–1063.

Chafetz, M. D. (2015). *Intellectual disability: Criminal and civil forensic issues (AACN workshop series).* New York, NY: Oxford University Press.

Chafetz, M. D., & Biondolillo, A. (2012). Validity issues in Atkins death cases. *The Clinical Neuropsychologist, 26,* 1358–1376.

Chafetz, M. D., & Dufrene, M. (2014). Malingering-by-proxy: Need for child protection and guidance for reporting. *Child Abuse & Neglect, 38,* 1755–1765.

Chafetz, M. D., & Underhill, J. (2013). Estimated costs of malingered disability. *Archives of Clinical Neuropsychology, 28*, 633–639.

Chafetz, M. D., Abrahams, J. P., & Kohlmaier, J. (2007). Malingering on the Social Security Disability Consultative Examination: A New Rating Scale. *Archives of Clinical Neuropsychology, 22*, 1–14.

Chafetz, M. D., Williams, M. A., Ben-Porath, Y. S., Bianchini, K. J., Boone, K. B., Kirkwood, M. W., Larrabee, G. J., & Ord, J. S. (2015). Official position of the American academy of clinical neuropsychology social security policy on validity testing: Guidance and recommendations for change. *The Clinical Neuropsychologist, 29*, 723–740.

Frazier, T. W., Frazier, A. R., Busch, R. M., Kerwood, M. A., & Demaree, H. A. (2008). Detection of simulated ADHD and reading disorder using symptom validity measures. *Archives of Clinical Neuropsychology, 23*, 501–509.

Green, P. (2004). *Manual for the Green's Medical Symptom Validity Test (MSVT)*. Edmonton: Green's Publishing.

Green, P. (2008). *Advanced interpretation (software)*. Edmonton: Green Publishing.

Green, P., & Astner, K. (1995). *The word memory test*. Edmonton: Neurobehavioural Associates.

Green, P., Flaro, L., & Courtney, J. (2009). Examining false positives on the word memory test in adults with mild traumatic brain injury. *Brain Injury, 23*, 741–750.

Greve, K. W., & Bianchini, K. (2006). Classification accuracy of the portland digit recognition test in traumatic brain injury: Results of a known-group analysis. *The Clinical Neuropsychologist, 20*, 816–830.

Greve, K. W., Bianchini, K., & Doane, B. M. (2006). Classification accuracy of the test of memory malingering in traumatic brain injury: Results of a known-group analysis. *Journal of Clinical and Experimental Neuropsychology, 28*, 1176–1190.

Haber, A. H., & Fichtenberg, N. L. (2006). Replication of the test of memory malingering (TOMM) in a traumatic brain injury and head trauma sample. *The Clinical Neuropsychologist, 20*, 524–532.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants. (2009). American academy of clinical neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist, 23*, 1093–1129.

Hoskins, L. L., Binder, L. M., Chaytor, N. S., Williamson, D. J., & Drane, D. L. (2010). Comparison of oral and computerized versions of the word memory test. *Archives of Clinical Neuropsychology, 25*, 591–600.

Howe, L. L. S., Anderson, A. M., Kaufman, D. A. S., Sachs, B. C., & Loring, D. W. (2007). Characterization of the medical symptom validity test in evaluation of clinically referred memory disorders clinic patients. *Archives of Clinical Neuropsychology, 22*, 753–761.

Kim, M. S., Boone, K. B., Victor, T., Marion, S. D., Amano, S., Cottingham, M. E., Ziegler, E. A., & Zeller, M. A. (2010). The warrington recognition memory test for words as a measure of response bias: Total score and response time cutoffs developed on "real world" credible and noncredible subjects. *Archives of Clinical Neuropsychology, 25*, 60–70.

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society, 18*, 1–7.

Larson, J. C. G., Flaro, L., Peterson, R. L., Connery, A. K., Baker, D. A., & Kirkwood, M. W. (2015). The medical symptom validity test measures effort not ability in children: A comparison between mild TBI and fetal alcohol spectrum disorder samples. *Archives of Clinical Neuropsychology, 30*, 192–199.

Loring, D. W., Larrabee, G. J., Lee, G. P., & Meador, K. J. (2007). Victoria symptom validity test performance in a heterogeneous clinical sample. *The Clinical Neuropsychologist, 21*, 522–531.

MacAllister, W. S., Nakhutina, L., Bender, H. A., Karantzoulis, S., & Carlson, C. (2009). Assessing effort during neuropsychological evaluation with the TOMM in children and adolescents with epilepsy. *Child Neuropsychology, 15*, 521–531.

Macciocchi, S. N., Seel, R. T., Alderson, A., & Godsall, R. (2006). Victoria symptom validity test performance in acute severe traumatic brain injury: Implications for test interpretation. *Archives of Clinical Neuropsychology, 21*, 395–404. doi:10.1016/j.acn.2006.06.003

Pankratz, L., & Erickson, R. D. (1990). Two views of malingering. *Clinical Neuropsychologist, 4*, 379–389.

Powell, M. R., Gfeller, J. D., Hendricks, B. L., & Sharland, M. (2004). Detecting symptom and test-coached simulators with the test of memory malingering. *Archives of Clinical Neuropsychology, 19*, 693–702.

144 (⊕) L. M. BINDER AND M. D. CHAFETZ

Schroeder, R. W., Peck, C. P., Buddin, W. H., Jr, Hargrave, D. D., Heinrichs, R. J., & Baade, L. E. (2012). The coin-in-the-hand test and dementia: More evidence for a screening test for neurocognitive symptom exaggeration. *Cognitive and Behavioral Neurology, 25*, 139–143.

Singhal, A., Green, P., Ashaye, K., Shankar, K., & Gill, D. (2009). High specificity of the medical symptom validity test in patients with very severe memory impairment. *Archives of Clinical Neuropsychology, 24*, 721–728.

Slick, D., Hopp, G., Strauss, E., & Thompson, G. B. (2005). *VSVT. Victoria symptom validity test. professional manual*. Lutz: Psychological Assessment Resources.

Slick, D. J., & Sherman, E. M. S. (2013). Differential diagnosis of malingering. In D. A. Carone & S. S. Bush (Eds.), *Mild traumatic brain injury: Symptom validity assessment and malingering* (pp. 57–72). New York, NY: Springer.

Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist (Neuropsychology, Development and Cognition: Section D), 13*, 545–561.

Slick, D. J., Tan, J. E., Strauss, E., Mateer, C. A., Harnadek, M., & Sherman, E. M. S. (2003). Victoria symptom validity test scores of patients with profound memory impairment: Nonlitigant case studies. *The Clinical Neuropsychologist (Neuropsychology, Development and Cognition: Section D), 17*, 390–394.

Social Security Administration (SSA). (2016a). *Annual statistical report on the social security disability insurance program, 2015*. Washington, DC: SSA Publication No. 13-11826.

Social Security Administration (SSA). (2016b). *SSI annual statistical report, 2015*. Baltimore, MD: Author.

Tombaugh, T. N. (1996). *TOMM. Test of memory malingering*. North Tonawanda, NY: Multi-Health Systems.

Vickery, C. D., Berry, D. T. R., Dearth, C. S., Vagnini, V. L., Baser, R. E., Cragar, D. E., & Orey, S. A. (2004). Head injury and the ability to feign neuropsychological deficits. *Archives of Clinical Neuropsychology, 19*, 37–48.

The Clinical Neuropsychologist
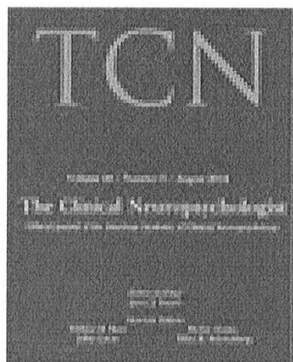
Intent to Fail: Significance Testing of Forced Choice Test Results

Laurence M. Binder[a], Glenn J. Larrabee[b] & Scott R. Millis[c]

[a] Independent practice, Portland, OR, USA

[b] Independent practice, Sarasota, FL, USA

[c] Department of Physical Medicine & Rehabilitation, Wayne State
University , Detroit, MI, USA
Published online: 11 Nov 2014.

PLEASE SCROLL DOWN FOR ARTICLE

Routledge
Taylor & Francis Group

Intent to Fail: Significance Testing of Forced Choice Test Results

Laurence M. Binder[1], Glenn J. Larrabee[2], and Scott R. Millis[3]

[1]Independent practice, Portland, OR, USA
[2]Independent practice, Sarasota, FL, USA
[3]Department of Physical Medicine & Rehabilitation, Wayne State University , Detroit, MI, USA

A score that is significantly below the level of chance on a forced choice (FC) performance validity test results from the deliberate production of wrong answers. In order to increase the power of significance testing of a below chance result on standardized FC tests with empirically derived cutoff scores, we recommend using one-tailed tests of significance and selecting probability levels greater than .05 (.20 for most standardized FC tests with empirically derived cutoff scores). Under certain circumstances, we also recommend combining scores from different sections of the same FC test and combining scores across different FC tests. These recommendations require modifications when applied to non-standardized FC tests that lack empirically derived cutoff scores or to FC tests with a non-random topographical distribution of correct and incorrect answers.

INTRODUCTION

Some widely cited neuropsychological criteria for diagnosis of malingering (Slick, Sherman, & Iverson, 1999) stated that significantly below chance performance on forced choice (FC) tests constituted evidence of definite malingering, when an external incentive was present. Consistent with the Slick et al. criteria, the American Academy of Clinical Neuropsychology Consensus Conference Statement (Heilbronner et al., 2009) asserted that significantly below chance performance on FC tests represented a "deliberate attempt to misrepresent one's own abilities" (p. 1103). A National Academy of Neuropsychology position paper (Bush et al., 2005) characterized similar performance with the phrase, "biased responding" (p. 422).

Significantly below chance performance on FC tests is common in some forensic neuropsychological settings. An early study using the Portland Digit Recognition Test (PDRT) showed that 17% of a sample of 102 examinees with financial incentives for claims of mild head injury performed significantly below chance, $p < .10$ (Binder, 1993). Subsequent studies used a p level of .05. With the Victoria Symptom Validity Test (VSVT), 6% performed worse than chance (Grote et al., 2000). Both the Word Memory Test (WMT) and the PDRT yielded significantly below chance performance in almost 6% of a large sample with financial incentives while the Test of Memory Malingering (TOMM) had a figure of 2.6% (Greve, Binder, & Bianchini, 2009). Using

---

GOVERNMENT EXHIBIT

4:21-CR-009-GCH
No. 123

GOVERNMENT EXHIBIT 123

2   **LAURENCE M. BINDER ET AL.**

various measures in a sample of criminal defendants, 22% performed below chance on at least one FC test (Ardolf, Denney, & Houston, 2007). Chafetz (2008), in a sample of Social Security Disability applicants, found that 12% scored below chance on any trial of the TOMM.

Most standardized FC tests involve two alternatives for each item. In a two-alternative FC test, the probability of randomly selecting or guessing the correct response, P, equals the probability of the incorrect response, Q, with P = Q = .50. FC tests have numerous two-alternative trials, for example, 36 for each of the easy and hard sections of the PDRT (Binder & Willis, 1991), and 50 for each trial block of the TOMM (Tombaugh, 1996). Scores on these performance validity tests (PVTs) can be analyzed as to whether or not they fall below the range expected for chance responding, using the normal approximation to the binomial, which defines the numerator as the number correct minus (N × P), with N representing the number of trials and P = 0.5 for a PVT with two alternatives for each item. The denominator is $\sqrt{NPQ}$ (Siegel, 1956). The probability of the resulting z-score is obtained from a z table.

A correction factor is utilized for the z approximation for less than 50 trials, because the normal distribution is for a continuous variable, while the binomial distribution involves a discrete variable. The correction factor of 0.5 is subtracted from the numerator when the number of errors exceeds 50% of the items. When there are 50 or more trials, regardless of the values of P and Q, the binomial approximation can be characterized as normal, and no correction is required. When there are between 36 and 49 trials, and when P = Q = .5 (a two alternative test), the approximation is very good; in these situations, the correction also is unnecessary (Runyon, Coleman, & Pittenger, 2000). In this paper, for computations of the binomial approximation of scores on two alternative FC tests, we will use the correction factor when the number of trials is less than 36.

Despite the diagnostic importance and the frequency of significant FC results, there is no rationale or consensus for selection of an appropriate level of probability to use for the assessment of statistical significance or for the related decision to use a one-tailed or two-tailed test. One issue is the choice of significance level. Although the VSVT authors (Slick, Hopp, Strauss, & Thompson, 2005) defined significance as .05, the PDRT author (Binder, 2007) recommended a p level of .10. A second issue is the decision to use one-tail or two-tailed tests, and a study presenting data on the frequency of below chance results using both one-tailed and two-tailed tests (Greve et al., 2009) implied lack of consensus. A third issue is the combining of scores from different sections of a test and combining scores from different FC tests. Combining scores enlarges the number of items undergoing significance testing. As the number of items increases, there is a reduction in the error percentage needed to obtain significance. Binder (2007) recommended testing of significance on the PDRT on the 36 Easy items, the 36 Hard items, and on the 72 Total items; a similar procedure is recommended for the VSVT (Slick et al., 2005). In the Manual for the TOMM, Tombaugh (1996) was silent on the issue of combining scores on the three trials of the test.

In this paper, we provide recommendations for testing of significance and interpretation of below chance scores on FC tests regarding the directionality of hypothesis testing, probability levels, and the combining of test scores.

DIRECTIONALITY OF HYPOTHESIS TESTING

Generally, neuropsychological testing involves detection of deficits; the poorer the examinee's performance, the greater his impairment. On a two alternative FC task, a complete absence of ability, or "zero ability", should result in a level of performance within the range of chance, or approximately 50% correct, whereas someone who is feigning impairment might be expected to perform worse than chance. A score that is significantly worse than chance demonstrates intentional avoidance of the correct choice, in an attempt to portray impairment.

Denney (1996) demonstrated that a group of community-dwelling volunteers who were unfamiliar with alleged criminal offenses demonstrated "zero ability" by performing at the chance level on FC testing of knowledge of the details of these crimes. Three criminal defendants suspected of malingering remote memory impairment in the context of evaluation for competency to proceed to trial performed at a significantly below chance level on correct identification of facts relevant to the alleged offenses, presented in two-alternative FC format.

A two-tailed test of significance is used for testing a hypothesis that one score is different in either direction than another score or for testing the hypothesis that a coin is biased in either direction for flips of heads or tails. A one-tailed test is appropriate for testing the hypothesis that a neurological group is impaired relative to the control group or that the coin is biased in one direction.

We recommend using a one-tailed test of significance when a score is lower than chance. Someone feigning impairment on a FC PVT will underperform by substituting wrong for right answers. A one-tailed test is appropriate for testing the hypothesis that a below chance score is significant, because the hypothesis is unidirectional – the performance either was or was not worse than chance. A one-tailed test yields a probability level that is half the probability of a two-tailed test. It is not meaningful to test for significance when a FC test score is better than chance. Better than chance scores on well-standardized PVTs should be compared with the relevant, empirically determined cutting scores.

PROBABILITY LEVELS

Testing of statistical significance in FC testing with an individual examinee whose scores are below the chance level tests a different hypothesis than a statistical test comparing group means. A below chance score is not compared with a group mean. In FC testing, it is hypothesized that a score from an examinee that is below chance is significantly worse than a score obtained by guessing the correct answer on each item. If the score is significantly worse than chance, then one logically concludes that the score was generated by *intentionally* providing wrong answers.

It has been suggested that the traditional .05 level of significance that is widely used for research on groups is inappropriate for assessing the significance of a below chance result on a FC test. Reynolds and Horton (2012) recommended the use of a probability level above .10 for FC tests. They reasoned that the .05 level of significance recommended by statistics pioneer R. A. Fisher for the comparison of experimental group means was not consistent with the mandate in the legal arena to provide testimony on a more probable than not basis. The specific p level to be employed was not

4   LAURENCE M. BINDER ET AL.

specified by Reynolds and Horton. In the PDRT Manual, Binder (2007), citing Reynolds (1998), also recommended consideration of probability levels higher than .10.

Forced choice PVT scores also are interpreted in terms of empirically developed cutoff scores. Widely used FC PVTs employ cutoff scores that are at or above chance. Cutoff scores are derived from studies of groups with well-defined neurological histories of brain disease or injury that do not have financial or other external incentives for suboptimal motivation, poor effort, or exaggeration of cognitive abnormalities. These empirically derived PVT cutoff scores are typically selected so that false positive identification of neurologically impaired subjects occurs at a rate of 10% or less; that is, 90% of patients with bona fide brain dysfunction can score above the cutoff and pass the PVT (Boone, 2007; Larrabee, 2008).

In consideration of empirically derived cutoff scores that are above chance, and the fact that the .05 level of probability is designed for testing the null hypothesis of differences of group means, we believe that the .05 level is too conservative for testing the hypothesis that a *standardized* FC test score is significantly worse than chance. We distinguish between *standardized* and *non-standardized* FC procedures. The latter are case specific, for example, testing knowledge of facts of a specific crime (Denney, 1996) or left–right discrimination. For *non-standardized* FC procedures that are case specific and that do not have empirically derived cutoff scores, we recommend a cutoff score associated with a normal approximation to the binomial probability of .10 or less, a value conforming to the false positive rates recommended for empirical cutoffs for Performance Validity Tests (Boone, 2007; Larrabee, 2008).

In order to determine an empirically justified p level for standardized FC tests, we examined data on the Warrington Recognition Memory Test–Words. These data were obtained from 136 participants in Michigan with claims of mild TBI, all of whom had financial incentives for their performance on neuropsychological testing (litigation group) and from 127 no-incentive outpatient participants with a history of TBI in the range of complicated mild to severe (TBI group). A score on the Recognition Memory Test–Words of less than 35 correct (70%) correctly classified 78% of the litigants and 89% of the TBI patients. No member of the TBI group scored worse than 23 (46%) correct. Analysis of the probability of significance of a score of 22 correct, the empirically derived cutoff with zero false positives (100% specificity) for this sample, using the z approximation, yielded $z = 0.8485$, $p = .2005$. The Recognition Memory Test cutoff of below 46% derived from the Michigan sample was supported by data in the literature on the same test. In a sample of 124 patients of mixed etiologies without external incentives for invalid performance who demonstrated good effort on other measures, the worst performance was 58% correct (Kim et al., 2010).

Therefore, we recommend a p level of .20 for significance testing of below chance results on *standardized* FC tests with empirically derived cutting scores that are at or above the level of 50% correct. To our knowledge, this includes all the well-known standardized FC tests with two alternatives for each item. Arguably, a less conservative p level can be justified.

Data obtained from other standardized FCTs also support the use of a p level of .20. To our knowledge, samples of people with brain dysfunction lacking incentives for poor performance do not include scores below chance on standardized FC tests with two alternatives for each item. For example, the TOMM Manual (Tombaugh, 1996) lists the cutting score for invalid performance as falling at or below 88% correct on Trial 2

or Delay. This value is significantly *better* than chance, $z = 5.37$, $p < .0001$. The poorest performance of any of the non-malingering clinical comparison group was 58% (a 70-year-old patient with Alzheimer's disease). The TBI group data in the TOMM Manual for Trial 2 shows that only one of 45 traumatic brain injury patients (a participant with three months loss of consciousness, a subdural hematoma, and a right hemiparesis) scored 44 correct (88%). The lowest score by a dementia patient (70 years old with Alzheimer's disease) was 29 correct (58%), and the lowest score in the cognitively impaired group, a patient with Korsakoff's syndrome, was 35 correct (70%).

Similar data were reported for the PDRT (Binder, 2007; Binder & Kelly, 1996). The original cutting scores (subsequently modified) were set to achieve 100% specificity in a traumatic brain injury sample with unequivocal evidence of brain dysfunction and no financial incentive. These scores were below 52.8% correct for the Easy items, below 50% for the Hard items, and below 54.2% for total items combined.

Examinees without external incentives for low scores also perform better than chance on the Victoria Symptom Validity Test (VSVT). Table 3 of the VSVT Manual (Slick et al., 2005) appears to show that none of a non-compensation-seeking patient group with seizure disorder and/or severe traumatic brain injury scored below 75% for the Easy or the Difficult items, or below 77% on the total correct; Table 4 of the Manual appears to show that none of a group of moderate to severe TBI performed below 83% for Easy or below 50% for Difficult items. The VSVT Manual data are consistent with data in the peer reviewed literature (Grote et al., 2000; Macciocchi, Seel, Alderson, & Godsall, 2006).

Classification statistics applied to either criterion group ("known group") or simulation research designs indicate that cutting scores exceeded by 100% of a clinical patient group are equivalent to definite performance invalidity. The Positive Predictive Power (PPP, or probability of performance invalidity) is equal to True Positives ÷ (True Positives + False Positives); when False Positives are 0, the PPP becomes 1.00 (Baldessarini, Finklestein, & Arana, 1983). The empirically derived cutting scores assure high specificity and allow the use of a p level greater than .05. The classification statistics based on the cutting scores, including sensitivity, specificity, and the base rate of performance invalidity, provide information about the chance of misclassification of performance on a PVT and the probability of invalid performance. The p level is not the probability of invalid performance.

We use scores on the TOMM to demonstrate the effect of increasing the p level on classification of scores as below chance. If one uses a one-tail test of significance and p of .05, a score of 19 or fewer correct is significant, $p = .046$, and a score of 20 is not significant, $p = .079$. A score of 21 correct, $p = .129$, is less than half the empirically-derived cutoff scores that were recommended by the test author and later researchers (Greve et al., 2006; Tombaugh, 1996). All members of Tombaugh's (1996) TBI group were correct on far more than half the items on all three trials. The lowest score of any of Tombaugh's clinical subjects on Trial 2 was the aforementioned subject from the dementia group who scored 29 correct. The probability level is the probability that a score was obtained by guessing or random responding. A score of 21 likely is the product of intentionally incorrect responding. Although some other FC tests are more difficult than the TOMM, as demonstrated by scores of well-motivated comparison groups with brain dysfunction, the same logic applies. Table 1 shows the number

6   LAURENCE M. BINDER ET AL.

Table 1. Cutoff scores of number correct for various levels of significance

	50 items	100 items	150 items	40 items	80 items	36 items	72 items	24 items*	48 items
.05	19	41	64	14	32	13	29	8	18
.10	20	42	67	15	34	14	30	9	19
.20	21	45	69	17	36	15	32	9	21

The listed scores are significantly below chance, for example, on a test with 50 items, a score of 20 correct, $p < .10$. The TOMM has a total of 150 items, including 50 items in each of 3 trials. The Word Choice Test and Warrington Recognition Memory Test–Words have 50 items. The Word Memory Test has a total of 80 items with two forced choice alternatives, including 40 Immediate Recognition items and 40 Delayed Recognition items. Scores on the Word Memory Test Consistency (CNS) items are dependent on scores on the Immediate and Delayed Recognition items, therefore, CNS scores should not be combined with Immediate and Delayed Recognition scores when testing for significance. The PDRT has a total of 72 items, including 36 Easy items and 36 Hard items. The VSVT has a total of 48 items, including 24 Easy items and 24 Hard items.

*Correction for continuity used in calculation of fewer than 36 items.

correct required for statistical significance at different probability levels across various numbers of test items.

COMBINING TEST SCORES

Statistical power increases with more test items. Several widely used forced choice tests have different parts whose scores can be combined. For instance, the VSVT has a total of 48 items divided into two parts, each with 24 items. The VSVT Manual (Slick et al., 2005) recommends computation of statistical significance on various sections of the test separately and in combination. A score of 8 out of 24 (33%) correct yields, $p = .078$ with a one-tail test, while a score of 16 out of 48 (33%) correct yields, $p = .011$. On the TOMM, with three sections, each containing 50 items, for 21 out of 50 (42%) correct, $p = .129$. If the same 42% correct is obtained on all three sections of the test, for 63 out of 150 items correct, $p = .026$, a dramatic difference in probability level.

We recommend combining scores on sections of forced choice tests such as the VSVT, TOMM, WMT, and PDRT when scores on each section are below 50% correct, in order to increase the number of items and the power of significance testing. A particular percentage correct on combined sections of these tests may reach a level of significance sufficient to describe performance as significantly worse than chance, while that same percentage correct on a single section of a test may not reach a sufficient level of probability.

Scores on different sections of tests can be combined only when the test items are independent. The concept of independence of items is straightforward. Consider the probability of picking a blue marble from a bag of seven marbles of assorted colors, including two blue marbles. The trials are not independent, because the probability changes after the selection of each marble, when sampling without replacement. Selection of a blue marble on the first pick lowers the probability of selecting a blue marble on the second pick from .286 to .167. In contrast, when tossing a fair coin, the trials are independent. Obtaining heads on the first toss does not alter the probability of obtaining

heads on the second coin toss. In the same way, the success of an examinee on the items on immediate recognition on the WMT, if he or she is guessing, is independent of success on the items on delayed recognition, if he or she is guessing. The independence of the *items* exists despite the fact that *scores* on immediate and delayed recognition are directly related.

The authors of FC tests recommend combining test scores. Paul Green, the author of the WMT and other forced choice measures, also combines scores on immediate and delayed recognition trials when testing for statistical significance (Green, 2008). The PDRT and VSVT authors also suggest combining scores (Binder, 2007; Slick et al., 2005).

Scores that are not independent should not be combined. Scores on the WMT consistency index are dependent on scores on immediate and delayed recognition and should not be combined with recognition memory scores for the purpose of testing of statistical significance.

Scores on different forced choice tests may be combined, in order to increase the power of significance testing. Combining scores is the proper procedure because the probability of guessing correctly on any item on one test is independent of the probability of guessing an item correctly on another test. For example, suppose an examinee on the WMT was correct on 18 items on immediate recognition and 18 items on delayed recognition. The resulting significance level for 36 of 80 items correct is $p = .187$. The same hypothetical examinee was correct on 20 of 48 items on the VSVT, $p = .125$. When the scores on the WMT and VSVT are combined, 56 correct of 128 items correct, $p = .079$.

Combining test scores across sections of one FC test or across different FC tests is recommended only to increase power when the scores to be combined are all below chance. If the examinee is significantly below chance on one FC test or section of a test and above chance on another FC test or section of a test, then we do not recommend erasing the below chance result by combining scores.

DISCUSSION

In summary, the likelihood of obtaining statistically significant results with below chance scores on FC tests will increase with the use of one-tailed hypothesis testing, probability levels above .05, and combining test scores within and between PVTs. These measures can be used in unison. We recommend that one-tailed tests and a p level of .20 be used for all below chance results on FC PVTs with empirically derived cutting scores that are at or above the level of chance. If below chance results are obtained from different sections of the same test, such as the three trials of the TOMM, then the scores can be combined. Scores also can be combined from different PVTs. These measures may be used in combination with one another, for example, p level of .20 and one-tailed testing and combining scores within and between PVTs.

There are FC tests that require a more conservative approach than the one we recommend. As Denney (1996) and others have shown, the examiner can construct a FC test using case-specific information. However, when using a case-specific FC task, cutoff scores based on the performance of well-motivated persons with brain

8   **LAURENCE M. BINDER ET AL.**

dysfunction usually are not available. Therefore, it is necessary to select a more conservative p level of .10 of than the .20 level selected for a well-standardized FC PVT with known cutoff scores.

Our recommendations also cannot be applied to PVTs if the probability of guessing right or wrong is related to the topographical location of the correct and incorrect items on the recognition probe. Most standardized FC tests were designed so that the stimulus probes contain an equal number of correct answers in particular locations (for FC tests with two alternatives, either left or right, or top or bottom). For example, on the PDRT and the TOMM, the correct answer is located at the top of the stimulus probe in half the items.

In contrast to some standardized PVTs, Green's Nonverbal Medical Symptom Validity and Medical Symptom Validity Tests contain stimulus probes with non-random locations of the correct answers (Paul Green, personal communication, May 6, 2014). We suggest a more conservative approach when applying significance testing to tests with non-random locations of the answers. This limitation does not affect the validity of cutting scores for these tests that were established by empirical research.

Implementation of our recommendations for using a one-tailed test and a p level of .20 should have a modest effect on the frequency of classification of PVT results as significantly below chance. For example, we examined data from Greve and Bianchini (2006) on PDRT Total scores in 56 participants classified as showing evidence of poor effort on other measures in their battery. Eleven of the 56 with poor effort on other measures had PDRT scores below chance at the level of $p < .20$ (one tailed) and nine had scores below chance at the level of $p < .10$ (one tailed). It could not be determined how many from this sample had scores significant at the .05 level, but it was clear that our suggested p level would not dramatically increase the frequency of classification. In a larger sample from the same population, 6% were worse than chance at the .05 level (Greve et al., 2009).

Deliberate exaggeration of impairment is common after possible mTBIs that are not expected to produce significant permanent impairment, rather than after moderate severe injuries where this exaggeration occurs less frequently (Binder, 1993; Greve & Bianchini, 2006). Motivation to show more impairment than truly exists can be documented by FC testing. If FC tests are failed using empirically derived cutoffs, then possible genuine deficits cannot be measured and there is no objective neuropsychological evidence of impairment. In the presence of significantly below chance results on FC testing, the examinee has demonstrated intent to provide wrong answers and to exaggerate impairment. Poor performances on cognitive tests may reflect intentional underperformance, while normal range scores may be underestimates of actual level of ability.

When there are invalid test data after a condition such as a severe TBI, the clinician can grossly estimate outcome using non-test data, for example, using return to work data as a function of initial time to follow commands (Dikmen et al., 1994). The likelihood of genuine impairment increases when there is a medical history that is likely to lead to permanent dysfunction, such as a severe TBI. In one case seen by the second author, the accuracy of the test data could not be relied upon due to evidence of performance invalidity, but the examinee's history of an admission Glasgow Coma Scale of 3, his 1-month time to follow commands, and his 3 months of post-traumatic amnesia showed injury severity characteristics associated with only 8% of the Dikmen et al. (1994) TBI subjects returning to work. In short, the results of FC testing must be placed

within the context of other test data and the examinee's history, and conclusions sometimes must be estimated from base-rate predictions derived from peer reviewed research.

The diagnosis of malingering may not be synonymous with the determination that all neuropsychological test performances are invalid. However, performance on a FC test that is significantly worse than chance provides definite evidence of malingering and deliberate response bias, when this occurs in the context of the presence of substantial external incentive (Heilbronner et al., 2009; Slick et al., 1999). We agree with Pankratz, who described significantly worse than chance performance as "the smoking gun of intent" (Pankratz & Erickson, 1990, p. 385).

ACKNOWLEDGEMENT

The authors thank the reviewers and the associate editor for their constructive comments.

DISCLOSURE

Laurence Binder, Glenn Larrabee, and Scott Millis perform medicolegal consultations. Laurence Binder is the author and publisher of the Portland Digit Recognition Test.

REFERENCES

Ardolf, B. R., Denney, R. L., & Houston, C. M. (2007). Base rates of negative response bias and malingered neurocognitive dysfunction among criminal defendants referred for neuropsychological evaluation. *The Clinical Neuropsychologist, 21*, 899–916.

Baldessarini, R. J., Finklestein, S., & Arana, G. W. (1983). The predictive power of diagnostic tests and the effect of prevalence of illness. *Archives of General Psychiatry, 40*, 569–573.

Binder, L. M. (1993). Assessment of malingering after mild head trauma with the Portland Digit Recognition Test. *Journal of Clinical and Experimental Neuropsychology, 15*, 170–183.

Binder, L. M. (2007). *Portland Digit Recognition Test Manual* (3rd ed). Beaverton, OR: Author.

Binder, L. M., & Kelly, M. P. (1996). Portland Digit Recognition Test performance by brain dysfunction patients without financial incentives. *Assessment, 3*, 403–409.

Binder, L. M., & Willis, S. C. (1991). Assessment of motivation after financially compensable minor head trauma. *Psychological Assessment, 3*, 175–181.

Boone, K. B. (2007). A reconsideration of the Slick et al. (1999) criteria for malingered neurocognitive dysfunction. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment. A neuropsychological perspective* (pp. 29–49). New York, NY: Guilford.

Bush, S. S., Ruff, R. M., Troster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., Reynolds, C. R., & Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity. *Archives of Clinical Neuropsychology, 20*, 419–426.

Chafetz, M. D. (2008). Malingering on the Social Security Disability consultative exam: Predictors and base rates. *The Clinical Neuropsychologist, 22*, 529–546.

Denney, R. L. (1996). Symptom validity testing of remote memory in a criminal forensic setting. *Archives of Clinical Neuropsychology, 11*, 589–603.

Dikmen, S. S., Temkin, N. R., Machamer, J. E., Holubkov, M. A., Fraser, R. T., & Winn, R. (1994). Employment following traumatic head injuries. *Archives of Neurology, 51*, 177–186.

10   **LAURENCE M. BINDER ET AL.**

Green, P. (2008). *Advanced Interpretation (software)*. Edmonton, Alberta: Green Publishing.

Greve, K. W., & Bianchini, K. J. (2006). Classification accuracy of the Portland Digit Recognition Test: Results of a known group analysis. *The Clinical Neuropsychologist, 20*, 816–830.

Greve, K. W., Bianchini, K. J., & Doane, B. M. (2006). Classification accuracy of the Test of Memory Malingering in traumatic brain injury: Results of a known-groups analysis. *Journal of Clinical and Experimental Neuropsychology, 28*, 1176–1190.

Greve, K. W., Binder, L. M., & Bianchini, K. J. (2009). Rates of below-chance performance in forced-choice symptom validity tests. *The Clinical Neuropsychologist, 23*, 534–544.

Grote, C. L., Kooker, E. K., Garron, D. C., Nyenhuis, D. L., Smith, C. A., & Mattingly, M. L. (2000). Performances of compensation seeking and non-compensation seeking samples on the Victoria Symptom Validity Test: Cross validation and extension of a standardization study. *Journal of Clinical and Experimental Neuropsychology, 22*, 709–719.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants. (2009). American academy of clinical neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist, 23*, 1093–1129.

Kim, M. S., Boone, K. B., Victor, T., Marion, S. D., Amano, S., Cottingham, M. E., Ziegler, E. A., & Zeller, M. A. (2010). The Warrington Recognition Memory Test for words as a measure of response bias: Total score and response time cutoffs developed on "real world" credible and noncredible subjects. *Archives of Clinical Neuropsychology, 25*, 60–70.

Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist, 22*, 666–679.

Macciocchi, S. N., Seel, R. T., Alderson, A., & Godsall, R. (2006). Victoria Symptom Validity Test performance in acute severe traumatic brain injury: Implications for test interpretation. *Archives of Clinical Neuropsychology, 21*, 395–404.

Pankratz, L., & Erickson, R. D. (1990). Two views of malingering. *The Clinical Neuropsychologist, 4*, 379–389.

Reynolds, C. R. (1998). Common sense, clinicians, and actuarialism in the detection of malingering during head injury litigation. In C. R. Reynolds (Ed.), *Detection of malingering during head injury litigation* (pp. 261–286). New York, NY: Plenum.

Reynolds, C. R., & Horton, Jr. A. M. (2012). Clinical acumen, common sense, and data-based decision making in the assessment of dissimulation during head injury litigation. In Reynolds, C. R., & Horton, Jr., A. M. (Eds.), *Detection of malingering during head injury litigation* (pp. 351–370). New York, NY: Springer.

Runyon, R. P., Coleman, K. A., & Pittenger, D. A. (2000). *Fundamentals of behavioral statistics* (9th ed.). New York, NY: McGraw-Hill.

Siegel, S. (1956). *Nonparemetric statistics for the behavioral sciences*. New York, NY: McGraw-Hill.

Slick, D., Hopp, G., Strauss, E., & Thompson, G. B. (2005). *VSVT. Victoria Symptom Validity Test. Professional Manual*. Lutz: Psychological Assessment Resources.

Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist, 13*, 545–561.

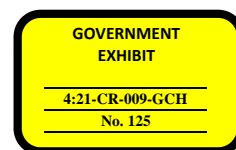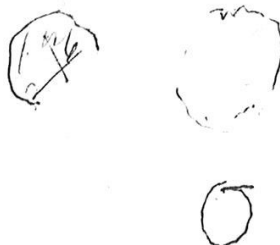Tombaugh, T. N. (1996). *Test of Memory Malingering*. Toronto: Multi-Health Systems Inc.

A B C
1 2 3
a b c
○ □ △
I II III
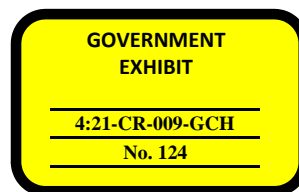
REY-15
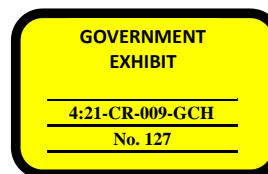10-26-21   Exposure

d ◇ II B 5 c

A III e ⬠ 2 —

☰ ☐ 6 1 F a

▱ 4 C b △ D

I E f = ○ 3

IGT2 Score Summary Table

Score	Raw score	Demographically corrected		U.S. Census-matched	
		T score	%ile	*T* score	%ile
NET TOTAL	16	49	46	51	54
NET 1	2	54	66	59	82
NET 2	0	45	31	46	34
NET 3	10	54	66	56	73
NET 4	6	50	50	51	54
NET 5	-2	43	24	43	24
Deck A′	16		>16		>16
Deck B′	26		>16		>16
Deck C′	22		>16		>16
Deck D′	36		>16		>16
Trials administered	100				
Total money	$-530				

Brockman
10/2/21

A
B
C
1
2
3 ← ?? enumeration

GOVERNMENT
EXHIBIT

4:21-CR-009-GCH
No. 128

GOVERNMENT
EXHIBIT
128
PENGAD-Bayonne, N.J.

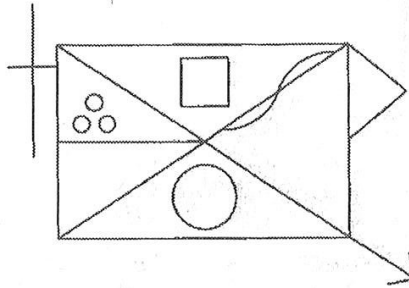## 12 Figure Recall

Say *Do you remember that figure that I had you copy? I want you to draw as much of it as you can remember now. If you remember a part, but you're not sure where it goes, put it anywhere. Try to draw as much of it as you can.*

Now, present the Figure Recall Drawing Page.

Scoring: 1 point for correctness and completeness (drawing), and 1 point for proper placement. See Appendix 1 in Stimulus Booklet B for complete scoring criteria and scoring examples.

Figure Recall Criteria
(Fold back for use.)

Item	Drawing (0 or 1)	Placement (0 or 1)	Score (0, 1, or 2)	Scoring Criteria
1. rectangle			2	Drawing: lines are unbroken and straight; angles 90 degrees; top/bottom lines 25% longer than sides Placement: not rotated more than 15 degrees
2. diagonal cross			2	Drawing: lines are unbroken and straight and should approximately bisect each other Placement: ends of lines should meet corners of the rectangle without significant overlap or measurable distance between the ends of the lines and the corners
3. horizontal line			0	Drawing: line is unbroken and straight; should not exceed 1/2 the length of the rectangle Placement: should bisect left side of the rectangle at approximately a right angle and intersect the diagonal cross
4. circle			0	Drawing: round, unbroken and closed; diameter should be approximately 1/4–1/3 height of rectangle Placement: placed in appropriate segment; not touching any other part of figure
5. 3 small circles			0	Drawing: round, unbroken and closed; equal size; triangular arrangement; not touching each other Placement: in appropriate segment; not touching figure; triangle formed not rotated more than 15 degrees
6. square			0	Drawing: must be closed; 90 degree angles; lines straight and unbroken; height is 1/4–1/3 height of rectangle Placement: in appropriate segment; not touching any other part of figure; not rotated more than 15 degrees
7. curving line			0	Drawing: 2 curved segments are approximately equal in length and symmetrical; correct direction of curves Placement: ends of line touch diagonal; do not touch corner of rectangle or intersection of diagonal lines
8. outside cross			0	Drawing: vertical line of the outside cross is parallel to side of rectangle; >1/2 the height of rectangle; horizontal line crosses vertical at 90 degree angle and is between 20–50% of length of vertical line Placement: horizontal line of outside cross touches rectangle higher than 2/3 the height of rectangle, but below top; does not penetrate the rectangle
9. triangle	0	\	1	Drawing: angle formed by 2 sides of triangle is between 60–100 degrees; sides are straight, unbroken and meet in a point; distance on vertical side of rectangle subsumed by triangle is approximately 50% of the height of vertical side Placement: no measurable distance between the top of the rectangle and the triangle
10. arrow			0	Drawing: straight and unbroken; lines forming arrow are approximately equal in length and not more than 1/3 length of staff Placement: must protrude from appropriate corner of rectangle such that staff appears to be continuation of diagonal cross
Total Score Range=0–20			5	

CONFIDENTIAL

RTBrockman_Medical_R

Figure Recall Drawing Page
(Fold back for use.)

RTBrockman_Medical_Records_0006290
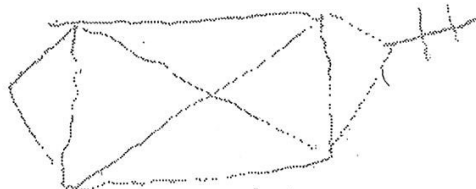
Core Memory Indices Across Time

	York 03/2019 WMS-4 LMVR II 53	York 12/2019 WMS-4 LMVR II 67	York 10/2020 WMS-4 LMVR II 57	Denney 5/2021 NAB Memory Index 58	Guilmette 7/2021 RBANS DMI 48	Guilmette 10/2021 RBANS DMI 52	Denney 10/2021 NAB Memory Index 58
Percentiles	0.1	1	0.2	0.26	0.1	0.1	0.26
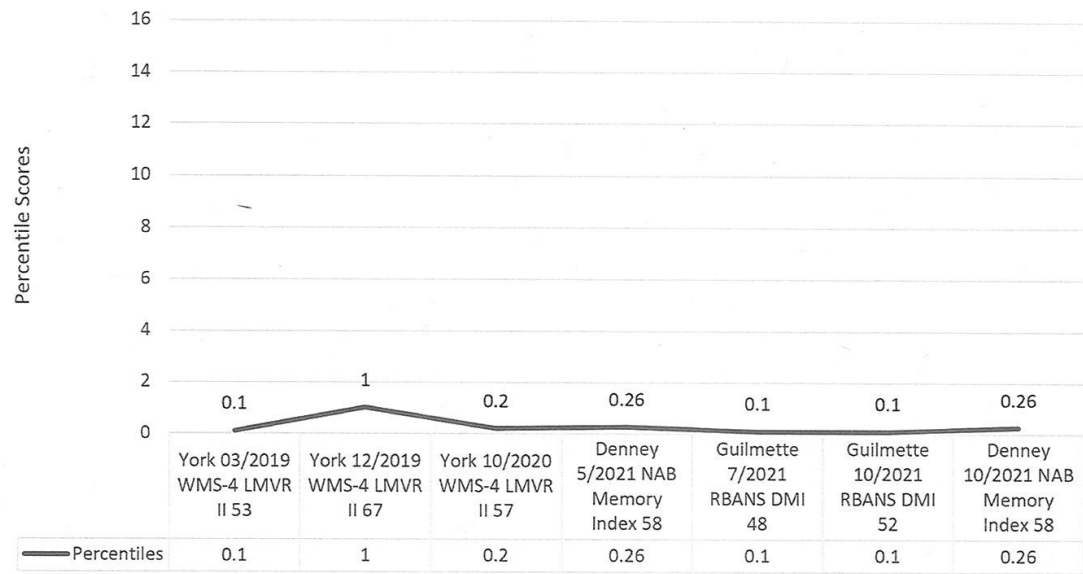
Don,
Evatt,

Earlier this spring I went to Washington DC to visit SAB and Miller & Chevalier to inquire after their retention of documents related to the 1992-1996 tax issues.

What I got from SAB was not great in volume - but had Brook Voght's files that he had brought with him from Miller & Chevalier relating to the various proposed structures for Hotrod. They gave all of these to me - and I have personally shredded them.

The volume from Miller & Chevalier was much, much greater - probably 15 banker boxes. These are the working files of Brook, George Hani, Holly Porter, Bob Catcher, and others.

I am going through these a page at a time because I wanted to shred personally anything that was super sensitive before turning them over to the outside document shredding and destruction service.

What I wanted to report to you both - is that the volume, detail, and organization of these files was dumbfounding.

For example - the organizational chart of the international entities that was built from Brook's notes had multiple copies - so far I have shredded half a dozen of these - and I still have 6 banker boxes to go through.

The face to face and telephone conversations that we had with M&C were immediately afterwards recounted, transcribed, and typed up - then copies circulated to everyone on the team - and ended up in everyone's working files. I can't tell you how many pages of this type of stuff I have shredded.

Obviously if attorney-client privilege ever broke down - that would an extremely unpleasant occurrence.

Therefore in any further encounters with the house that require the usage of outside tax counsel we need to:

-obtain upfront agreement from the attorneys that we get to review their files - including those of all attorneys that work on the case
-that we get to dictate what gets retained and what does not (like background info on international structures)
-be very circumspect as to what we tell them verbally - as it will all be reduced to writing and kept in the file
-be thoughtful about any piece of paper that they get, as it will for sure be kept in the file of every attorney that works on the case
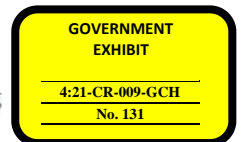
In general, I was super-impressed with the organization and the thoroughness of M&C. It also turns out the the workhorse in this deal (judging by file volume) was Holly Porter.

Bob

From:	Permit
To:	"Redfish"
Subject:	RE:
Date:	Monday, June 19, 2017 4:34:00 PM

Evatt,

I agree with your concerns. The position I have taken so far with friends has caused no further questions – however the crew is a different story.

I think the best course of action is for you to tell Tom that I am a minority owner with certain exclusive rights of usage. This and any further details are confidential for him only.

.

As far as the crew is concerned with the relationship, I should be described as "an owner", similar to the relationship I enjoyed on the Trevia.

Bob

From: Redfish [mailto:redfish@hannah.com]
Sent: Sunday, June 18, 2017 7:06 PM
To: 'Permit'
Subject:

Bob,

I need to raise with you a matter which is causing me concern.

Today, Tom Henshilwood asked to see me privately to get "clarity" about a few things.

The main thing he wanted clarity about was ownership of the Albula. It took me by surprise that he would raise the issue, but we have seen a great deal of speculation about the owner of the boat. In fact, one of the occupants of the Savannah – the grey, modern boat at the marina - asked me on Friday evening if Bob Brockman was on the Albula. He would like to have given you a tour of the Savannah. His name is Bas Nederpelt, the commercial director of Feadship, the Dutch shipyard. Savannah was opened up to people with the potential to order a boat from Feadship. You are already tagged as the owner of the Albula in the superyacht community.

I don't mind that Tom put me on the spot, though the only thing I could say is the following:

-        I represent the Foundation that is the majority owner.
-        That Foundation is engaged in supporting marine research conducted by leading universities.
-        Bob Brockman is a minority owner with rights of exclusive use
-        Other than the philanthropic aspect, I have zero interest in boats and much prefer to be on land, which is why I don't sleep on the Albula and leave you and Dorothy to decide many matters to do with the boat.

What Tom's question tells me is that you are already accepted as the sole and absolute owner of the Albula. The only way the director of Feadship could have your name is through crew members of the Albula and crews of the different boats talking to each other.

I am doing what I can to dispel this idea where it can cause you harm. When asked about the ownership, as I was by Tom, I explained the marine research objective.

I try to use the boat in ways that would suggest that you are not the sole owner, but this is disruptive to my work and family time. Frankly, my efforts are lame since you are – as far as Tom is concerned – the chief decision maker while I am an infrequent user who has never slept on the boat and avoids going out on it. Sophie has been carrying a lot of this for me.

Tom even asked me today what the stewardesses should say to Dorothy about me using the boat. He said that he would prefer to not have the girls lie about me using the boat. This was absolute confirmation that you are the owner. Clearly Tom believed I had lied about the ownership of the boat. Tom also told me today that Dorothy believed the Albula was returning to Florida as soon as you left Bermuda. I guess he was gently accusing me of keeping the boat here and using it without permission from you and Dorothy.

The problem with all of this is that it elevates your profile in a way that I struggle to address. If questioned by a regulator, I would have no trouble explaining Mountain Queen or the Henke properties. I am confident I can protect you and I say that knowing that it was Robert Smith buying property through a trust that has triggered the DOJ problems.

The boat is different. It is not in my hands to control the loose lips of itinerant crew members. I cannot argue that the acquisition of the boat is an investment that will enhance in value in the same way as I can point to the enhanced asset value of the Aspen property.

You asked me to use the boat before and after your use to justify the Foundation bearing the cost of positioning the boat. This would also reinforce that you are not the sole or majority owner thereby explaining why the Foundation contributed about 90% of the acquisition cost. If I have Tom and the crew effectively saying that I am using the boat without your permission, then obviously we have failed.

The crux of the matter is this: unless I can find a way to allow you to openly declare the boat is yours and yours alone, I need to present you as a minority owner. Right now the only way I can think of doing the former is for you to buy all shares in the Cayman company and pay the annual operating costs. That is not acceptable. I'll keep working on a better way to do that.

In order to provide sufficient cover for you, it would greatly help me if you could get Tom and the crew to accept what I told him today. I hate to ask, but I think it is also important, if at all possible, for Dorothy to understand better the situation with the ownership issues and the need for the Fisheries Research Foundation to cover costs and the positioning of the boat for you.

I would very much like to find a way for you to have full use of the boat year round with all costs

covered by the Foundation. As matters stand today, the risk to you and me is high because of Robert Smith. I do believe we are being monitored – I certainly am. The boat has placed an even bigger target on us.

Evatt

TO DO LIST – TANGARRA

revised December 17, 2011

VISTA EQUITY FUND`

-made $600M initial committement to VEPFIV

-also consider debt fund that uses leverage and buys debt of Vista portfolio companies

PROJECTS – NEW – Investment Management Advisory Company

-it is time to proceed on this project by contacting Debevoise & Plimpton so that it can start small and start getting some age on it

-this should be an LLC owned directly by RTB that provides investment advice to Point

-this entity could also provide management services to portfolio companies allowing further income that is tax deductible to the portfolio companies

-the LLC should be US based with a year-to-year contract with Point such that at some point a new entity that included RTB II could be formed to take over the business

-investments should be made directly by Point as this is an already established entity through which SSHLTD has historically made investments – which have provided the AEBCT with a further liability shield

-fee percentage should be industry standard and may be different based upon the investment mix

-fee is adjusted annually

-amount of fee is regulated by how much capital is committed

-consider Debevoise for drafting the investment management company contract

-this entity only provides recommendations to Point, it does not act for Point so as to avoid any FBAR requirements

CONFIDENTIAL - ATTORNEY/CLIENT WORK PRODUCT   **page** 1 of 12
12/28/2010 9:11 AM
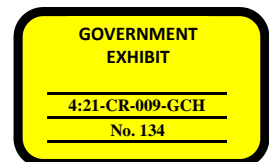
PROJECTS – OTHER

-build combined structure charts for the AEBCT structure (send me a complete set of charts)

-republish structure charts after the reorganization of the three non-AEBCT structures are done

-keep a running spreadsheet on private equity commitments

-build a complete description of all the 2011 organizational changes

-build for the years 2010-2011 a log of significant transactions – and continue to update this log forever – so that it is easy to see how asset balances move around

-create master document by entity of all bank accounts, signatories, and contact persons within the banks – republish when any changes occur – **add section for comments – such as relationship details – key people**

-due to Bermuda laws that enable assets to be frozen with essentially no right to effectively protest thru the courts, all significant bank balances should be elsewhere where the rule of law and the courts is still in place – probably Switzerland – or maybe Panama

-this should include all information necessary to be able to log on to these accounts to verify balances

-remaining sets of externally kept books are – Spartacus, Augustus (Dan Jenkins), SSHLTD (Jim Collins), Point (now Evatt with Dan Voth for backup)

-need to find a way that someone else runs Point in order to satisfy the auditors

PROJECTS - TRUSTEES & TRUST PROTECTORS

-establish Graham Wood as Trust Protector personally for the AEBCT, he then resigns and Acquitaine becomes the corporate Trust Protector where Graham Wood is Director

-document telephone conversation with Trevor Lloyd prior to his death where he
        resigns and concurs with your suggestion as Graham Wood personally as
        successor – provide an original wet-ink signed copy of this memo to Bob

-have Graham Wood personally resign and do a formal deed appointing his corporation signed in wet ink

-for backup purposes, secure a wet-ink signed letter of resignation and appointment of a new Trust Protector with date and appointee left blank – send this to Bob

-secure a digital signature from Graham Wood along with a wet-ink signed authorization to use it, provide this signature to Bob along with an original of this authorization

-prepare an Adobe version of the formal deed with date and appointee left blank ready for use with the digital signature – send a copy of this to Bob

-Heather now works part-time – where she continues to handle St. John's, AEBCT, SSHLTD, signatory on Point accounts, books for SSHLTD, Rome Investments, and the local Bermuda scholarship activities

-visit Al Thorpe to establish record of trustee visiting one of the directors of UCSH

-annual review of Carlos Kepke files

-annual trust meeting to be set at a new date and time – perhaps mid-semester break where RTB II can come to Bermuda (and Al Deaton as well)

-cultivate Mark Patterson as a future Trust Protector, perhaps a golf trip to Bermuda

-have RTB II visit Bermuda to meet Heather, BCB, BNTB

-begin to consider how to provide backup to Evatt – qualifications are:
        -accounting background
        -not a US citizen
        -capable of acquiring a Bermuda work permit of at least 10 years
        -might discover someone thru the scholarship program

PRIVATE MUTUAL FUND ADMINISTRATION

-liquidate Arboria - $3.1M recovered, rest is tied up on the backend of Madoff

-liquidate Founding Partners - ongoing

CURRENT INVESTMENTS – this section is starting to be re-opened due to Reynolds refinancing, Ventyx sale, and the improving environment

-Investment Policy for Mutual Fund Investments
     -Class 4 – 0% - money market – hold for emergency transfers
     -Class 3 – 50-60% low risk equity
     -Class 2 – 20-25% medium risk equity
     -Class 1 – 15-20% higher risk equity

-TradeTrakker
     Furnish new copies of portfolio whenever buy/sell events take place

-monitor current investment performance
     -30 day deposits getting .25%

-follow for any potential recovery
     -Bristol – write-off – continue to follow for recovery – partial distribution of $1.7M
     received – some potential for minor recovery $140-150K

-complete losses
     -STIR – liquidated at $5M loss in total
     -Core Digital – write-off
     -Rosefaire FCS 10% interest $1.8M
     -A/R from Gate –

NEW INVESTMENTS

-periodically think about asset allocation

-receipt of cash from sale of Reynolds is unlikely for the next 2 years according to current thinking – then the amount will be around $2B

-VFF $50M commitment

-Sunquest 2nd lien debt $29.487M

-Vision Solutions 2nd lien debt $26M

-Applied 2nd lien debt $20.524M

-new investment with VEPFIV, debt fund

-if management buyout of Reynolds does not take place and an outright sale takes place, start a private equity type operation focusing on software companies like Vista does using former Reynolds personnel

-set up an investment management firm that advises Point (and if advisable – SSHLTD)

POTENTIAL CHARITABLE DONATIONS
    -Stuart Yudofsky - $15M more

CHARITABLE DONATIONS
    -Rice University – Physics Building - $20M
    -Centre College $19.5M committment
    -Lance Gould $4.5M
    -Stuart Yudofsky - $10M

DOCUMENTS

-preparation of doomsday groups of documents for former Edge and Cabot structures once their re-organization is done

RECORDS

-St. John's has two boxes – **send me the inventory on these – so as to make a final decision as to whether to re-locate their contents to Houston**

TRIPS

-trips to Danville (Oct 19, 2012 is likely date of Residential Commons dedication), Houston, and Rice

-trip to Alaska – Labor Day Sept 3, 2012 – arrive in Houston by Sunday Sept 2th at the latest

-Nevis – meet Ernie Dover Q3 2012

-BVI with Don Q1 2012

-USA – quarterly

EVATT

-working on a larger family

STRATEGIC LONG-TERM

-Bob & Dorothy redo wills for US estate setting US testamentary trust for care of Robert II

-Evatt sends to Bob <u>a new version</u> of the academic articles on the Testamentary Power of Appointment

-in the event of premature demise of RTB, offshore world (other than AEBCT and its affiliates) should be held for RTB II to eventually manage (Dorothy will be involved in the AEBCT structure)

-the long term vision is as follows:

      -RTB hopefully lives a long, healthy life
      -after Reynolds RTB engages in private equity investments much like Vista
      -former Reynolds key personnel join this organization – or alternatively
          participate in a LBO/Preferred Stock bailout/MBO
      -RTBII finishes his education in a technical area and also gets an MBA
      -RTBII eventually finds the right lady and starts a family (and stays in Houston)
      -RTBII joins RTB in private equity operation – or investment advisory operation
      -the private equity operation earns fees for managing AEBCT funds at industry
          standard rates providing a living for RTB and RTBII
      -the AEBCT continues to engage in charitable giving in the areas of higher
          education and medical research
      -the corpus of AEBCT is eventually rolled over to AEBGCT or one of the PBB
          trusts

-the disaster scenario where something untoward happens to RTB is as follows:

      -a US-based trusted person holds the ownership papers of the trust protector
          Corporations
      -this US-based person (Al Deaton) endeavors to guide RTBII in the long term
          Vision laid out above

      -there is a backup person (Robert Burnett) to the US-based trusted person
          referred to above

ENTITY STRUCTURE

Cabot Structure – Reorganization

-create a new charitable trust – purpose medical research
-Choice and Barrier divend shares in Cabot up to Inverness and Shetland Trusts
-Inverness and Shetland Trusts gift their Cabot shares up to Aberdeen Trust
-Aberdeen Trust transfers these Cabot shares to Addington Trading
-Performance dividends its cash up to Endurance Trust
-Endurance Trust gifts this cash to Aberdeen which places the cash into Addington
-Aberdeen gifts the shares in Addington to the new charitable trust
-Addington has all the shares of Cabot, plus cash, plus share in Edge liquid and ill-liquid
-Lineage II Trust – holds some shares of Edge – what to do?

Edge Structure – Reorganization

-The Benevolent
      -Software Ltd pays off as much as it can on its loan to Plattoon
      -Plattoon's cash from the liquid portion of Edge goes to Regency
      -Advertising Services – as soon as one remaining account is closed

-create new charitable trust – for educational purposes

-Plattoon (Legend Trust) gifts its ill-liquid Edge to Legend (Heritage Trust)
-Cascade (Service Trust II) gifts its liquid and ill-liquid Edge to Legend (Heritage Trust)

-reduce the number of entities - or at least focus on getting rid of inactive entities

COMMUNICATIONS/PC ISSUES

-find a good double entry accounting system with the following requirements
  -GL detail records must contain date, name/address account number, memo number, document number, control number, and description field as well as G/L account number and transaction amount
  -traditional journals
    -cash receipts
    -cash disbursements
    -general journal
  -no money entries can be made except by journal entry
  -lots of historical data in the GL account records
  -maintains detail history forever on selected accounts
  -ability to suppress print when detail entries with the same control number zero balance
  -possessing standard G/L reports as well as P&L
  -ability to map all fields including historical monthly, quarterly, and yearly fields into Excel
  -Sage PeachTree is one to start the search with

-no usage of logic.bm   no unencrypted usage of houstonfishingservice.com except in emergency when lambdaprime.org is down

-document boot-up process for the email server

-send a copy of the boot CD to Bob

-transition all third parties to use this email address, leaving only personal email on logic.com

-LPG generator with one-week's supply of fuel

-build a 3rd email laptop – lower in priority than other "hardening" processes

-fire extinguishers for new house

-standards for files saved as part of the document database
  -relevant emails should be saved in .msg format
  -spreadsheets should be saved in .xls format
  -quarter and year-end reports should be saved in .pdf format
  -emails and reports from accounting systems should be saved as part of the regular data archiving process

CONFIDENTIAL - ATTORNEY/CLIENT WORK PRODUCT   **page** 9 of 12
12/28/2010 9:11 AM

STANDARD MEETING AGENDA

-review cash balances
-review investment performance
-review structure documentation
-discuss status of each structure
-review status of files **– need new inventory of files now that inactive files are
     segregated**
-review status of PC systems and software

PROJECTS – NEW – Aircraft Charter Company – now postponed indefinitely

-seek bids for the cash sale of the RJ to independent 3rd parties, then set the price of the
RJ a little above that to compensate for the "orderly sale" delay

-sell RJ to a new US-based C-corp – with preferably RTB owning 1% so he can fly at
cost

-the sale must be of a depreciated asset so as to achieve the EBITDA gain intended

-investigate the potential tax consequences of the sale and subsequent purchase by a
1% owner – does this impact the ability of the new entity to depreciate the aircraft?

-after this change, the RJ is chartered to Reynolds and other third parties at the going
$5500/hour rate – and to RTB at the owner's $2500 rate

ENTITIES IN WIND-DOWN MODE – BUT STILL NOT GONE

-shut down Software BV – in liquidation process – final complete in 2017

DANGLING THREADS

-CITGO - used to be in BVI, but now Heritage documents that they have are in Georgetown, will be retained forever - used to own CTL, Master, Kojak, Micro-Mainframe, Peters - left there is correspondence with Don, trust deed copies, being returned to BVI hopefully – status is that they cannot find these files

-Bank of Bermuda - there will be some left over internal memos written by bank officers in their personal file at least and there are microfilm copies of incoming mail

-Carlos' correspondence, billing, multiple destination, and computer files (gone)

-files of Ken, Robert, Craig - Mexico

-Brook Voght=s files - whatever is left on their network server and its backups, same for what is at his new law firm

George Hani=s files - same issue as Brook

Ben's Jiltec and SFL records - initial reply was 10 years - find out what is the situation on Ben=s correspondence files (maybe destroyed by hurricane)

Butterfield Cayman - computerized accounting records of everything - CTL, Jiltec, SFL, Kojak

Baring Bros. - TIL audit reports, Carlos' correspondence, trust variance documents

VP-Bank copying of structure files and who knows what else – Glen Godfrey

VP-Bank bank records

ATU – files left behind

Malta – DBA, Providian – they will not destroy the copy of anything they ever sent out to us

Edge involvement in VEFII

Don's paper files that still may be about

Don's miscellaneous computer files

Peter Poole's office

CONFIDENTIAL - ATTORNEY/CLIENT WORK PRODUCT   **page** 11 of 12
12/28/2010 9:11 AM

While I was in Houston a question came up as to the Edge connection to VEFII.

I can report that by way of Sale and Purchase agreement dated August 24, 2004, $154,582,366 was paid by Point in consideration for Edge's interest in VEFII.

What we cannot tell is how this and Edge's other connection with VEFII was recorded by Vista.

It might be a worthwhile project in the future to review Vista's records and request destruction to the extent that we can.

-On May 23, 2011 Evatt (once again) sent and email through the etamine@logic.bm email account. This message contained an attachment of an Excel spreadsheet listing of all the original documents kept in Houston. This email will be kept in perpetuity in the backup tapes of logic.bm. This is the worst breach of security ever to happen to date.

From: Redfish PC
To: "Permit"
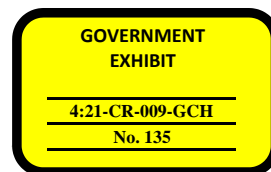**Date:**   Sunday, October 6, 2013 6:32:25 PM

Bob,

I'll be in New York from tomorrow morning until Wednesday afternoon.

I'll be deleting all email addresses from my iphone before going to the Bermuda airport and will set them up again in the taxi going to the hotel. I should be out of email contact until about midday New York time.

Evatt

From:	Permit
To:	"Redfish PC"
Subject:	RE:
Date:	Monday, May 5, 2014 5:44:00 AM

Evatt,

Please send to the office.

I agree with your thoughts here.

Bob

From: Redfish PC [mailto:redfish@proventusconstans.com]
Sent: Sunday, May 04, 2014 12:25 PM
To: Permit
Subject: Re:

Bob,

Will do. Do you want this sent to your home address rather than the office?

I have had a thought about chartering the vessel. Does it always need to be you? To demonstrate a variety of charterers could we have occasions where Al Deaton or Robert Smith charter the vessel?   There might be safe ways we could reimburse them - particularly Robert as the transaction could be entirely offshore.

Evatt

On 4 May 2014, at 14:42, "Permit" <permit@proventusconstans.com> wrote:

> Evatt,
>
> Please print off this attachment, sign all three, and then FED-X them back to me.
>
> By basing my comp plan off of EBITDA rather than pre-tax income, this will give me sufficient additional funds to be able to charter Turmoil for a block of 10 weeks per year at $150K per week plus fuel.
> That would cover almost completely the $4500/day that it costs to keep her.
>
> As you can see from the Whereas section, this is an opportune time to make this change due to the 8 key officers being put on an EBITDA based plan. Not to mention the logic that the value of Reynolds is no longer overshadowed by debt – and will be valued by EBITDA completely going forward.
>
> Bob
>
> <Employment Agreement for RTB 140101.doc>
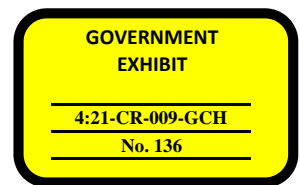
From:	Permit
To:	"redfish@proventusconstans.com"
Subject:	RE:
Date:	Monday, May 13, 2013 5:52:00 PM

Evatt,

I was also under the impression that the advice was to disclaim the special testamentary power of appointment – and also prevent any successor to that power.

That way when the service gets ahold of the trust document – they are not able to say that the power of appointment gives rise to the assertion of control over the trust – which necessitates registration under FATCA.

Bob

From: Redfish PC [mailto:redfish@proventusconstans.com]
Sent: Thursday, May 09, 2013 3:09 PM
**To:** permit@proventusconstans.com
Subject:

Bob,

Attached is a draft of the exercise of the special testamentary power of appointment. The idea here is that you would exercise the power in a way that would enshrine the independence of the trustee and ensure that no-one else could exercise the power.

As discussed, this document could be executed by you and held by you and/or Al to be produced if we ever face enquiries from a hostile party.

If not needed or laws change in some advantageous way, then the document could be destroyed and a different document prepared.

Evatt.
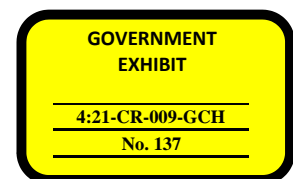
From:	Permit
To:	"Redfish"
Subject:	RE:
Date:	Thursday, April 2, 2015 1:56:00 PM

Evatt,

This is the type of charity that helps just one person with no multiplier effect.

My recommendation is to stick to charitable giving that has multiplier effects such as education and medical research.

Bob

From: Redfish [mailto:redfish@hannah.com]
Sent: Monday, February 09, 2015 2:48 PM
To: 'Permit'
Subject:

Bob,

I am sorry to ask you, but would consider a charitable donation to put a 16-year-old girl through boarding school for the next two years.

We've been contacted about a young girl we know in Bermuda, Kyla Lowe, who is now homeless.

Kyla's father is your typical Jamaican – she has never met him and he has provided nothing. Her mother is little better than useless. Unbeknown to Kyla, her mother stopped paying school fees three years ago and Kyla was finally kicked out of school. She has been at a community college this school year doing a High School equivalency course meant for adults with full time employment.

Kyla's grandmother provided housing, but she died a few months ago and the housing is now gone.

We have taken Kyla in but she needs to be at a proper school. She also needs to be away from her useless mother.

She is a good person who has worked hard all her life in trying circumstances and would probably do very well in the right conditions. We thought the school where Olivia and Keziah go would provide very well for her. She has never had stability in housing etc. Olivia would keep an eye on her and help her settle in.

Is she someone we can support? Going forward, whenever Kyla is back in Bermuda, she'll live with us. We'll look after all those needs.
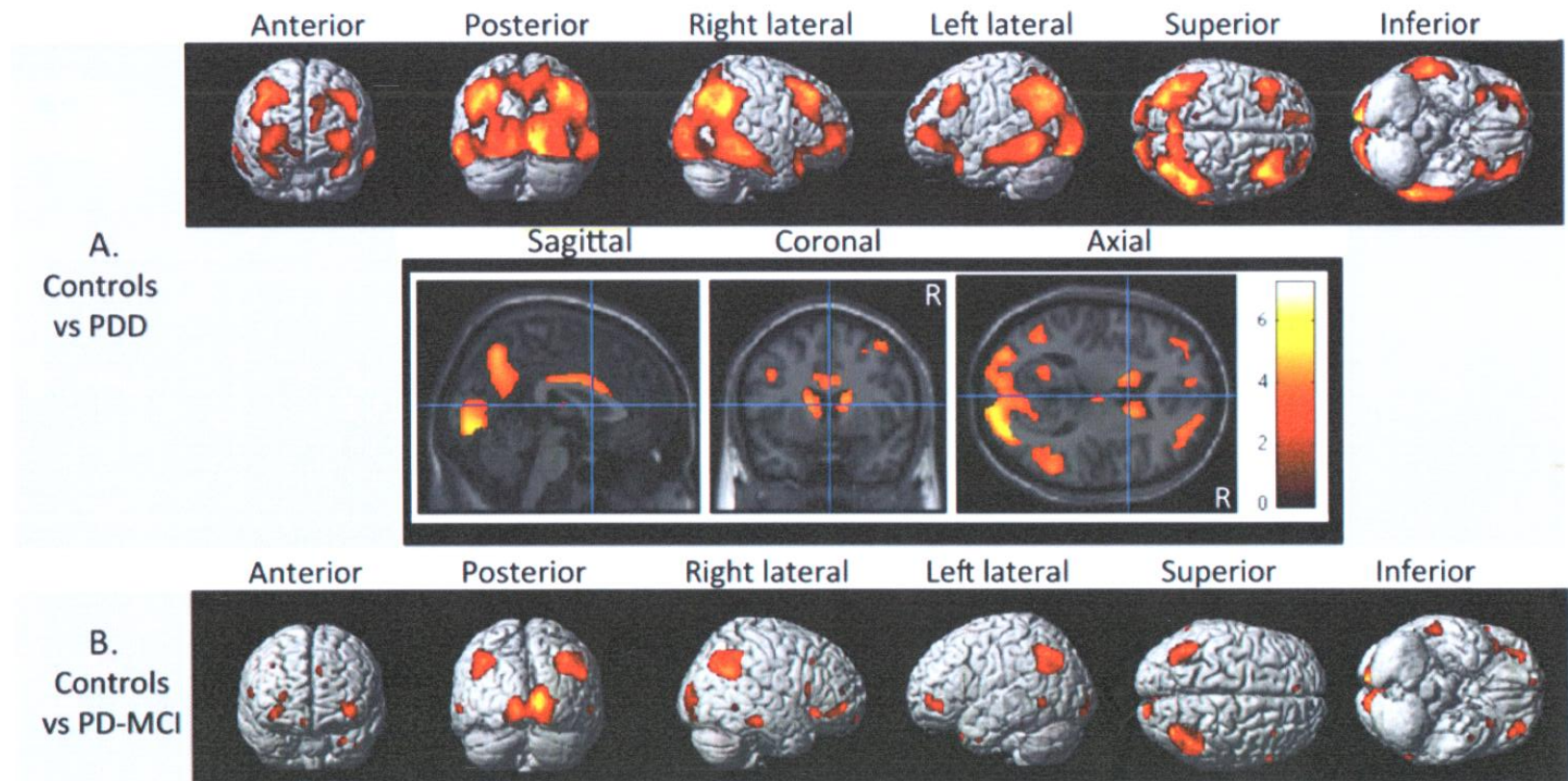
Evatt
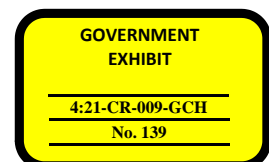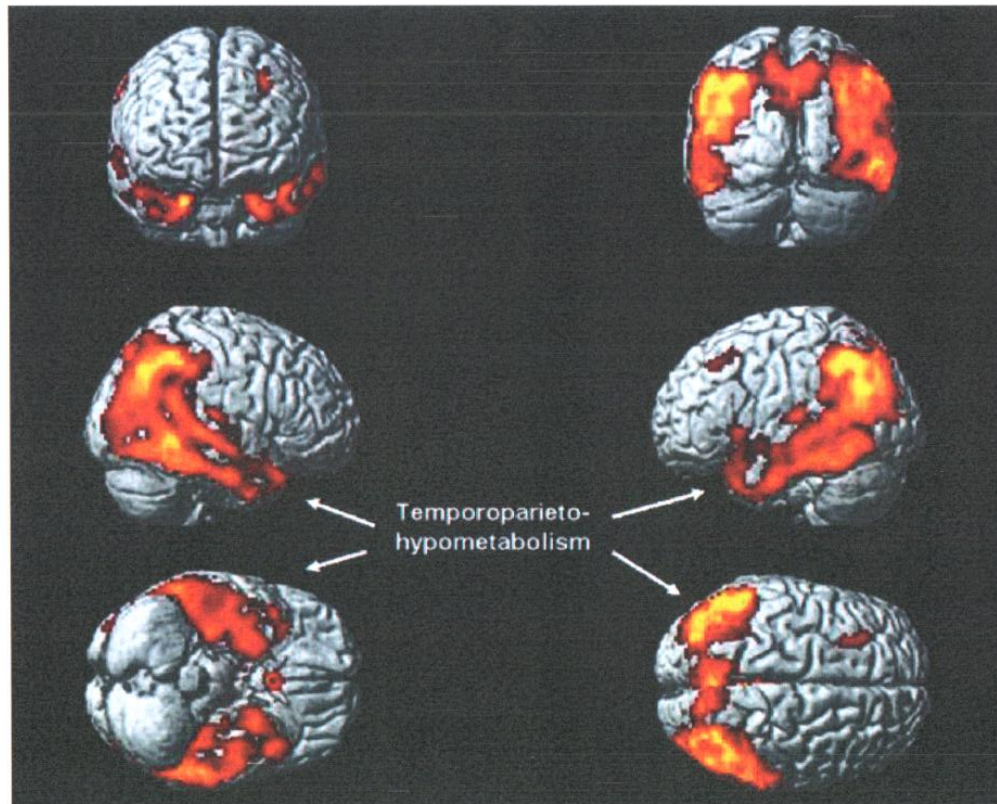
Figure 3. Reduction in temporoparietal glucose metabolism comparing 12 subjects with Alzheimer disease vs 8 controls in SPM (p < 0.001).

A. Controls vs PDD

B. Controls vs PD-MCI

Fig. 2 Regions with reduced metabolism comparing PDD patients and PD-MCI patients with respect to control subjects: *A* PDD<controls, *B* PD-MCI<controls ($p < 0.05$ FDR corrected; age and GDS score as covariates)